09/834,477

```
                    GenCore version 4.5
             Copyright (c) 1993 - 1998  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         May 13, 1999, 09:34:43 ; Search time 736.46 Seconds
                                         (without alignments)
                                         3458.727 Million cell updates/sec

Title:          US-08-704-178-1
Perfect score:  712
Sequence:       1 ATGGACCTGCAGCTGACCCA..........GGACCACGGTCACCGTCTCC 712

Scoring table:  IDENTITY_NUC

Searched:       808301 seqs, 1788773984 residues

Database :      GenEmbl:*
                1:  gb_ba1:*
                2:  gb_ba2:*
                3:  gb_in:*
                4:  gb_om:*
                5:  gb_ov:*
                6:  gb_pat:*
                7:  gb_ph:*
                8:  gb_pl1:*
                9:  gb_pl2:*
                10:  gb_pr1:*
                11:  gb_pr2:*
                12:  gb_pr3:*
                13:  gb_ro:*
                14:  gb_st:*
                15:  gb_sy:*
                16:  gb_un:*
                17:  gb_vi:*
                18:  gb_htg:*
                19:  em_ba:*
                20:  em_fun:*
                21:  em_hum1:*
                22:  em_hum2:*
                23:  em_in:*
                24:  em_om:*
                25:  em_or:*
                26:  em_ov:*
                27:  em_pat:*
                28:  em_ph:*
                29:  em_pl:*
                30:  em_ro:*
                31:  em_sy:*
                32:  em_un:*
                33:  em_vi:*
                34:  em_htg:*
                35:  em_sts:*
                36:  gb_ba1:*
                37:  gb_ba2:*
                38:  gb_pl1:*
                39:  gb_pl2:*
                40:  gb_pr1:*
                41:  gb_pr2:*
                42:  gb_pr3:*
                43:  gb_sts:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                   SUMMARIES
                  %
Result          Query
  No.   Score   Match Length DB  ID                               Description
-------------------------------------------------------------------------------
```

```
     1   699.6   98.3    711  6  I32406                    I32406 Sequence 1
     2   699.6   98.3    711 15  S39590                     S39590 anti-erbB2
     3     458   64.3    720  6  E13599                    E13599 DNA encodin
     4     434   61.0    732  6  I31035                    I31035 Sequence 1
     5     434   61.0    732 27  E10361                     E10361 cDNA encodi
     6     422   59.3    720  6  I32407                    I32407 Sequence 2
     7   356.2   50.0    916  6  I45604                    I45604 Sequence 29
     8   348.4   48.9    729  6  I31036                    I31036 Sequence 3
     9   348.4   48.9    729 27  E10362                     E10362 cDNA encodi
    10     338   47.5    744 15  SYNDA44X                   L17037 Synthetic s
    11   320.6   45.0    738 15  SYNAR3IGA                  M34000 Synthetic m
    12     319   44.8    797  6  AR012822                  AR012822 Sequence
    13   317.8   44.6    803  6  AR012824                  AR012824 Sequence
    14   315.8   44.4    735  6  E13598                    E13598 DNA encodin
    15   313.2   44.0    741  6  I08678                    I08678 Sequence 10
    16     311   43.7    831 13  MMZ95480                   Z95480 Mus musculu
    17   303.2   42.6    753  6  I08679                    I08679 Sequence 12
    18   296.6   41.7    819  6  A18692                    A18692 Synthetic n
    19   290.4   40.8    336 13  MMU37860                   U37860 Mus musculu
    20   285.8   40.1    819  6  A52549                    A52549 Sequence 2
    21   280.6   39.4    354 13  S78745                     S78745 Ig VH-anti-
    22   279.6   39.3    366 13  AF081541                   AF081541 Mus muscu
    23   279.6   39.3    366 13  MMVAR604                   Z68169 m.musculus
    24   277.6   39.0    355 13  S74055                     S74055 IgG VH IIA=
    25   277.6   39.0    351 13  MMU25678                   U25678 Mus musculu
    26   277.6   39.0    354 13  MMAHCVR2                   X90877 M.musculus
    27   274.2   38.5    414  6  I31960                    I31960 Sequence 84
    28   274.2   38.5    414  6  I78572                    I78572 Sequence 84
    29   274.2   38.5    414  6  I78627                    I78627 Sequence 84
    30   274.2   38.5    723 13  AF035617                   AF035617 Mus muscu
    31   274.2   38.5    357 13  MUSCMVG                    M95944 Mouse mRNA
    32     274   38.5    711  6  I08677                    I08677 Sequence 8
    33   272.8   38.3    421 13  S51851                     S51851 M104E myelo
    34   272.8   38.3    822 13  MMZ95479                   Z95479 Mus musculu
    35   272.4   38.3    755  6  AR012823                  AR012823 Sequence
    36   272.4   38.3    321 13  MMIG013                    X58586 Mouse hybri
    37   272.2   38.2    315  6  A36642                    A36642 Sequence 2
    38   272.2   38.2    315  6  I51657                    I51657 Sequence 3
    39   271.4   38.1    318 13  AF056217                   AF056217 Mus muscu
    40   270.6   38.0   9209  6  AR000007                  AR000007 Sequence
    41   270.2   37.9   1370 13  MUSIGKAAR                  K01641 Mouse Ig ka
    42   269.6   37.9    411 13  MMU88682                   U88682 Mus musculu
    43   269.4   37.8    312 13  S48339                     S48339 Ig V kappa
    44   269.2   37.8    345 13  MUSIGKVA2A                 M97864 Mouse hybri
    45   268.8   37.8    384  6  AR000010                  AR000010 Sequence


                                   ALIGNMENTS


RESULT   1
I32406
LOCUS       I32406        711 bp    DNA             PAT       07-JAN-1997
DEFINITION  Sequence 1 from patent US 5587458.
ACCESSION   I32406
NID         g1823197
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 711)
  AUTHORS   King,C.Richter, Kasprzyk,P.G. and Bird,R.E.
  TITLE     Anti-erbB-2 antibodies, combinations thereof, and therapeutic and
            diagnostic uses thereof
  JOURNAL   Patent: US 5587458-A 1 24-DEC-1996;
FEATURES             Location/Qualifiers
     source          1. .711
                     /organism="unknown"
BASE COUNT      175 a    182 c    190 g    164 t
ORIGIN


  Query Match             98.3%;  Score 699.6;  DB 6;  Length 711;
  Best Local Similarity   99.7%;  Pred. No. 2.6e-205;
  Matches  710;  Conservative    1;  Mismatches    0;  Indels    1;  Gaps    1;

Qy        1 ATGGACCTGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTA 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 ATGGACCTGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGT- 60

Qy       61 CACAATGACTTGCAGGGCCACCCCAAGTGTAAGTTACATGCACTGGTATCAGCAGAAGCC 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       60 CACAATGACTTGCAGGGCCACCCCAAGTGTAAGTTACATGCACTGGTATCAGCAGAAGCC 119

Qy      121 AGGATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGTCCCTGC 180
            ||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||
Db      120 AGGATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTGGCTTCTGGAGTCCCTGC 179

Qy      181 TCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGAGGC 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      180 TCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGAGGC 239

Qy      241 TGAAGATGCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGGAGG 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      240 TGAAGATGCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGGAGG 299

Qy      301 GGGGTCCAAGCTGGAAATAAAAGGTTCTACCTCTGGTTCTGGTAAATCTTCTGAAGGTAA 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      300 GGGGTCCAAGCTGGAAATAAAAGGTTCTACCTCTGGTTCTGGTAAATCTTCTGAAGGTAA 359

Qy      361 AGGTGTGCAGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTCAATGAAGAT 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      360 AGGTGTGCAGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTCAATGAAGAT 419

Qy      421 ATCCTGCAAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAAGCAGAG 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      420 ATCCTGCAAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAAGCAGAG 479

Qy      481 CCATGGAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATACTAACTA 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      480 CCATGGAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATACTAACTA 539

Qy      541 CAACCAGAAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCACAGCCTA 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      540 CAACCAGAAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCACAGCCTA 599

Qy      601 CATGGAGCTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGGAGGGT 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      600 CATGGAGCTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGGAGGGT 659

Qy      661 TACGGACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCC 712
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      660 TACGGACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCC 711

RESULT    2
S39590
LOCUS       S39590        711 bp             SYN       10-FEB-1993
DEFINITION  anti-erbB2 imnmunotoxin antigen binding region [mice, Other
            Synthetic Recombinant Partial, 711 nt].
ACCESSION   S39590
NID         g251113
KEYWORDS    .
SOURCE      Mus sp.
  ORGANISM  Mus sp.
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 711)
  AUTHORS   Batra,J.K., Kasprzyk,P.G., Bird,R.E., Pastan,I. and King,C.R.
  TITLE     Recombinant anti-erbB2 immunotoxins containing Pseudomonas exotoxin
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 89 (13), 5867-5871 (1992)
  MEDLINE   92335198
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 108547] from the original journal article.
            This sequence comes from Fig. 1.
FEATURES             Location/Qualifiers
     source          1. .711
                     /organism="Mus sp."
                     /db_xref="taxon:10095"
     gene            1. .711
                     /partial
                     /gene="anti-erbB2 imnmunotoxin antigen binding region"
     CDS             1. .711
                     /partial
                     /gene="anti-erbB2 imnmunotoxin antigen binding region"
                     /note="This sequence comes from Fig. 1."
                     /codon_start=1
                     /db_xref="PID:g251114"
                     /translation="MDLQLTQSPAILSASPGEKVTMTCRATPSVSYMHWYQQKPGSSP
                     KPWIYTTSNLASGVPARFSGGGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGS
                     KLEIKGSTSGSGKSSEGKGVQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQS
                     HGKNLEWIGLINPYNGDTNYNQKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCAR
                     RVTDWYFDVWGAGTTVTVS"
BASE COUNT      175 a    182 c    190 g    164 t
ORIGIN

  Query Match             98.3%;  Score 699.6;  DB 15;  Length 711;
  Best Local Similarity   99.7%;  Pred. No. 2.6e-205;
  Matches  710;  Conservative    1;  Mismatches    0;  Indels    1;  Gaps    1;

Qy        1 ATGGACCTGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTA 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 ATGGACCTGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGT- 60

Qy       61 CACAATGACTTGCAGGGCCACCCCAAGTGTAAGTTACATGCACTGGTATCAGCAGAAGCC 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       60 CACAATGACTTGCAGGGCCACCCCAAGTGTAAGTTACATGCACTGGTATCAGCAGAAGCC 119

Qy      121 AGGATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGTCCCTGC 180
            ||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||
Db      120 AGGATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTGGCTTCTGGAGTCCCTGC 179

Qy      181 TCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGAGGC 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      180 TCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGAGGC 239

Qy      241 TGAAGATGCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGGAGG 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      240 TGAAGATGCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGGAGG 299

Qy      301 GGGGTCCAAGCTGGAAATAAAAGGTTCTACCTCTGGTTCTGGTAAATCTTCTGAAGGTAA 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      300 GGGGTCCAAGCTGGAAATAAAAGGTTCTACCTCTGGTTCTGGTAAATCTTCTGAAGGTAA 359

Qy      361 AGGTGTGCAGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTCAATGAAGAT 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      360 AGGTGTGCAGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTCAATGAAGAT 419

Qy      421 ATCCTGCAAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAAGCAGAG 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      420 ATCCTGCAAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAAGCAGAG 479

Qy      481 CCATGGAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATACTAACTA 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      480 CCATGGAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATACTAACTA 539

Qy      541 CAACCAGAAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCACAGCCTA 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      540 CAACCAGAAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCACAGCCTA 599

Qy      601 CATGGAGCTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGGAGGGT 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      600 CATGGAGCTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGGAGGGT 659

Qy      661 TACGGACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCC 712
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      660 TACGGACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCC 711
```

```
RESULT    3
E13599
LOCUS       E13599        720 bp    DNA             PAT       27-APR-1998
DEFINITION  DNA encoding a signal chain Fv antibody of anti-gp130 antibody
            GPX7.
ACCESSION   E13599
NID         g3252401
KEYWORDS    JP 1997220092-A/2.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 720)
  AUTHORS   Ekida,T., Yasukawa,K., Imanaka,T. and Takagi,M. .
  TITLE     PRODUCTION OF SINGLE-STRAND FV ANTIBODY
  JOURNAL   Patent: JP 1997220092-A 2 26-AUG-1997;
            TOSOH CORP
COMMENT     OS   None
            OC   Artificial sequences.
            PN   JP 1997220092-A/2
            PD   26-AUG-1997
            PF   15-FEB-1996 JP 1996027622
            PI   EKIDA TEIJI, YASUKAWA KIYOSHI, IMANAKA TADAYUKI, PI   TAKAGI
            MASAHIRO
            PC   C12N15/09,C12N1/21,C12P21/08,(C12N1/21,C12R1:19),(C12P21/08,
            PC   C12R1:19);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .720
            FT                   /organism='Artificial sequences'.
FEATURES             Location/Qualifiers
     source          1. .720
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      169 a    193 c    190 g    168 t
ORIGIN

  Query Match             64.3%;  Score 458;  DB 6;  Length 720;
  Best Local Similarity   80.2%;  Pred. No. 8.8e-131;
  Matches  576;  Conservative    1;  Mismatches  131;  Indels   10;  Gaps    3;

Qy        4 GACCTGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTACAC 63
            || ||   ||| |||||||||||||||| | |||||||||||||||||||||||||| |||
Db        1 GAGCTCGTGCTCACCCAGTCTCCAGCATTGATATCTGCATCTCCAGGGGAGAAGGT-CAC 59

Qy       64 AATGACTTGCAGGGCCACCCCAAGTGTAAGTTACATGCACTGGTATCAGCAGAAGCCAGG 123
             ||||| ||||  | || | ||||||| | || |||| | ||||| ||||||||||||||
Db       60 CATGACCTGCAATGTCAGCTCAAGTGTTACTTCCATGTATTGGTACCAGCAGAAGCCAGG 119

Qy      124 ATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGTCCCTGCTCG 183
            ||||||||||||||| |||||||||  ||||||||||||:|||||||||||||||| ||||
Db      120 ATCCTCCCCCAAACCCTGGATTTATCTCACATCCAACCTGGCTTCTGGAGTCCCTCCTCG 179

Qy      184 CTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGAGGCTGA 243
            || ||||||| |||||||||||||||| |||||||||||||| |||||||  ||||||||||
Db      180 CTCCAGTGGCAGTGGGTCTGGGACCACTTACTCTCTCACAATCAGCAGCTTGGAGGCTGA 239

Qy      244 AGATGCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGGAGGGGG 303
            ||||||||||||||||||||||||||||||||||  || |||| | |||||||||| |  ||
Db      240 AGATGCTGCCACTTATTACTGCCAGCAGTGGAGTACTAACCCGCTCACGTTCGGTGCTGG 299

Qy      304 GTCCAAGCTGGAAATAAAAGGTTCTACCTCTGGTTCTGGTAAATCTTCTGAAGGTAAA-- 362
            | ||||||| ||  |    ||   |  ||| ||   ||||   ||   ||  ||   |  
Db      300 GACCAAGCTCGAGCTGGGTGGCGGTGGCTCGGGCGGTGGTGGGTCGGGTGGCGGCGGATC 359

Qy      362 -GGTGTGCAGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTCAATGAAGAT 420
             |  || ||  ||||| |||| ||||||||  ||||||||||||| | |||| |||||||
Db      360 GGACGTCCAATTGCAGCAGTCTGGACCTGAACTGGTGAAGCCTGGGGCTTCAGTGAAGAT 419

Qy      421 ATCCTGCAAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAAGCAGAG 480
            | |||||||| |||| || ||| ||||||||| | ||| |||| ||||||||||||||||
Db      420 ACCCTGCAAGGCTTCAGGATACACATTCACTGACTACAACATGGACTGGGTGAAGCAGAG 479

Qy      481 CCATGGAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATACTAACTA 540
            ||||||||||| ||||||||||||||||  ||||||| || | | ||||| || || |||
Db      480 CCATGGAAAGAGCCTTGAGTGGATTGGAGATATTAATTCTCATAGTGGTGGTATTATCTA 539

Qy      541 CAACCAGAAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCACAGCCTA 600
            |||||| |||||||||| |||||||| ||| |||||||| ||||| ||||||||||||||
Db      540 CAACCAAAAGTTCAAGGACAAGGCCATATTGACTGTAGATAAGTCCTCCAGCACAGCCTA 599

Qy      601 CATGGAGCTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGGA---- 657
            ||||||||||| ||| |||||||||||||||| |||||||||||||| ||||||||| |    
Db      600 CATGGAGCTCCGCAGCCTGACATCTGAGGACACTGCAGTCTATTATTGTGCAAGAACCTA 659

Qy      657 --GGGTTACGGACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCC 712
                ||    | |   | ||| ||   |||||||  ||| |||||  |||| ||||||
Db      660 CTATGTGTACGGCCACTTCTTTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCC 717


RESULT    4
I31035
LOCUS       I31035        732 bp    DNA             PAT       20-DEC-1996
DEFINITION  Sequence 1 from patent US 5582826.
ACCESSION   I31035
NID         g1821826
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 732)
  AUTHORS   Shimamura,T., Hamuro,J., Nakazawa,H., Kanayama,Y., Sugamura,K. and
            Takeshita,T.
  TITLE     Monoclonal antibodies which bind the gamma chain of human
            interleukin-2 receptor
  JOURNAL   Patent: US 5582826-A 1 10-DEC-1996;
FEATURES             Location/Qualifiers
     source          1. .732
                     /organism="unknown"
BASE COUNT      180 a    204 c    182 g    166 t
ORIGIN

  Query Match             61.0%;  Score 434;  DB 6;  Length 732;
  Best Local Similarity   78.6%;  Pred. No. 2.2e-123;
  Matches  574;  Conservative    1;  Mismatches  136;  Indels   19;  Gaps    4;

Qy        1 ATGGACCTGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTA 60
            |||||| | | ||||||||||||||||| |||| |||||||||||||| ||||||| | ||  
Db        1 ATGGACATCCTGCTGACCCAGTCTCCATCAATCATGTCTGCATCTCTAGGGGA-ACGGGT 59

Qy       61 CACAATGACTTGCAGGGCCACCCCAAGTGTA------AGTTACATGCACTGGTATCAGCA 114
            ||| ||||| ||||  |||| | ||||||||      |||||| |||||||||| |||||
Db       60 CACCATGACCTGCACTGCCAGCTCAAGTGTAAGTTCCAGTTACTTGCACTGGTACCAGCA 119

Qy      115 GAAGCCAGGATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGT 174
            |||||||||||||||||||||||  |||||||||| ||||||||||||:|||||||||||
Db      120 GAAGCCAGGATCCTCCCCCAAACTCTGGATTTATAGCACATCCAACCTGGCTTCTGGAGT 179

Qy      175 CCCTGCTCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGT 234
            ||| ||||||||||||||| ||||||||||||||||||||||||||||| |||||||  |
Db      180 CCCAGCTCGCTTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGCAT 239

Qy      235 GGAGGCTGAAGATGCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTT 294
            ||||||||||||||||||||||||||||||||| ||||    ||||  ||| | ||||||
Db      240 GGAGGCTGAAGATGCTGCCACTTATTACTGCCACCAGTATCATCGTTCCCCGCTCACGTT 299

Qy      295 CGGAGGGGGGTCCAAGCTGGAAATAAAAGGTTCTA---CCTCTGGTTCTGGTAAATCTTC 351
            ||| |  ||| ||||||||||  | ||||     |   |||| || |||||       ||
Db      300 CGGTGCTGGGACCAAGCTGGAGCTCAAAGTCGACAAATCCTCAGGATCTGGCTCCGAATC 359
```

```
Qy     352 TGAAGGTAAAGGTGTGCAGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTC 411
             || | |     ||  | ||  ||||||| ||| ||||| |||||| |||||||| |||
Db     360 CAAAAGCACGCAGGTCAAACTCGAGGAGTCTGGATCTGAGCTGGTGAGGCCTGGAGCTTC 419

Qy     412 AATGAAGATATCCTGCAAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGT 471
           | ||||| | ||||||||| ||||||| ||| |||||||  || ||   ||| |||||||
Db     420 AGTGAAGCTGTCCTGCAAGGCTTCTGGCTACACATTCACCAGCTACTGGATGCACTGGGT 479

Qy     472 GAAGCAGAGCCATGGAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGA 531
           ||||||||| |||||| |   ||||||||||||||||  |||| |||||   | ||||
Db     480 GAAGCAGAGGCATGGACAAGGCCTTGAGTGGATTGGAAATATTTATCCTGGTAGTGGTAG 539

Qy     532 TACTAACTACAACCAGAAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAG 591
           |||||||||| |  ||||||||||| ||||||||||| | ||||||||||  || |||||
Db     540 TACTAACTACGATGAGAAGTTCAAGAGCAAGGCCACACTGACTGTAGACACATCCTCCAG 599

Qy     592 CACAGCCTACATGGAGCTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGC 651
           ||||||||||||| | |||  ||| |||||||||||||||||||| |||||||||||| |
Db     600 CACAGCCTACATGCACCTCAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTAC 659

Qy     652 AAGGAGGGTTACGGACTGGTACTTCGATG---------TCTGGGGCGCAGGGACCACGGT 702
           ||| ||      | |||||  || | |||          ||||||   ||| ||| | ||
Db     660 AAGAAGCAGCCGGAACTGGGTCTACTATGCTATGGACTACTGGGGTCAAGGAACCTCAGT 719

Qy     703 CACCGTCTCC 712
           ||||||||||
Db     720 CACCGTCTCC 729


RESULT   5
E10361
ID   E10361     standard; RNA; ROD; 732 BP.
XX
AC   E10361;
XX
NI   d1108698
XX
DT   08-OCT-1997 (Rel. 52, Created)
DT   08-OCT-1997 (Rel. 52, Last updated, Version 1)
XX
DE   cDNA encoding a monoclonal antibody against human Interleukin 2
DE   receptor gamma chain.
XX
KW   JP 1995313188-A/1.
XX
OS   Mus sp.
OC   Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Rodentia;
OC   Sciurognathi; Muridae; Murinae; Mus.
XX
RN   [1]
RP   1-732
RA   Shimamura T., Hamuro J., Nakazawa H., Kanayama Y., Sugamura K.,
RA   Takeshita T.;
RT   "IMMUNOSUPPRESSIVE AGENT";
RL   Patent number JP 1995313188-A/1, 05-DEC-1995.
RL   AJINOMOTO CO INC, SUGAMURA KAZUO,.
XX
CC   OS   Mus sp. (mouse)
CC   PN   JP 1995313188-A/1
CC   PD   05-DEC-1995
CC   PF   21-APR-1994 JP 1994082836
CC   PR   21-APR-1993 JP 93P    94491, 07-MAR-1994 JP 94P    36065
CC   PI   SHIMAMURA TOSHIAKI, HAMURO JUNJI, NAKAZAWA HARUMI,
CC   PI   KANAYAMA YUKA,
CC   PI   SUGAMURA KAZUO, TAKESHITA TOSHIICHI
CC   PC   C12P21/08,A61K39/395,A61K39/395,C12N1/21,C12N5/20,
CC   PC   C12N15/13//C12N15/06,
CC   PC   (C12P21/08,C12R1:19),(C12P21/08,C12R1:91),(C12N1/21,C12R1:19);
CC   CC   strandedness: Double;
CC   CC   topology: Linear;
CC   FH   Key             Location/Qualifiers
CC   FH
```

```
CC   FT   source          1. .732
CC   FT                   /organism="Mus sp."
CC   FT                   /cell_type="hybridoma"
CC   FT                   /cell_line="GP-2"
CC   FT   mat_peptide     1. .732
CC   FT                   /product="anti-IL-2 receptor gamma chain"
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .732
FT                   /organism="Mus sp."
XX
SQ   Sequence 732 BP; 180 A; 204 C; 182 G; 166 T; 0 other;

  Query Match             61.0%;  Score 434;  DB 27;  Length 732;
  Best Local Similarity   78.6%;  Pred. No. 2.2e-123;
  Matches  574;  Conservative    1;  Mismatches  136;  Indels   19;  Gaps    4;

Qy       1 ATGGACCTGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTA 60
           |||||| | | |||||||||||||||| ||||| |||||||||||| |||||| | ||
Db       1 ATGGACATCCTGCTGACCCAGTCTCCATCAATCATGTCTGCATCTCTAGGGGA-ACGGGT 59

Qy      61 CACAATGACTTGCAGGGCCACCCCAAGTGTA------AGTTACATGCACTGGTATCAGCA 114
           ||| ||||| ||||  |||| | ||||||||      |||||| |||||||||| |||||
Db      60 CACCATGACCTGCACTGCCAGCTCAAGTGTAAGTTCCAGTTACTTGCACTGGTACCAGCA 119

Qy     115 GAAGCCAGGATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGT 174
           |||||||||||||||||||||||  |||||||||| ||||||||||||:|||||||||||
Db     120 GAAGCCAGGATCCTCCCCCAAACTCTGGATTTATAGCACATCCAACCTGGCTTCTGGAGT 179

Qy     175 CCCTGCTCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGT 234
           ||| ||||||||||||||| ||||||||||||||||||||||||||||| |||||||  |
Db     180 CCCAGCTCGCTTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGCAT 239

Qy     235 GGAGGCTGAAGATGCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTT 294
           ||||||||||||||||||||||||||||||||| ||||    ||||  ||| | ||||||
Db     240 GGAGGCTGAAGATGCTGCCACTTATTACTGCCACCAGTATCATCGTTCCCCGCTCACGTT 299

Qy     295 CGGAGGGGGGTCCAAGCTGGAAATAAAAGGTTCTA---CCTCTGGTTCTGGTAAATCTTC 351
           ||| |  ||| ||||||||||  | ||||     |   |||| || |||||       ||
Db     300 CGGTGCTGGGACCAAGCTGGAGCTCAAAGTCGACAAATCCTCAGGATCTGGCTCCGAATC 359

Qy     352 TGAAGGTAAAGGTGTGCAGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTC 411
             || | |     ||  | ||  ||||||| ||| ||||| |||||| |||||||| |||
Db     360 CAAAAGCACGCAGGTCAAACTCGAGGAGTCTGGATCTGAGCTGGTGAGGCCTGGAGCTTC 419

Qy     412 AATGAAGATATCCTGCAAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGT 471
           | ||||| | ||||||||| ||||||| ||| |||||||  || ||   ||| |||||||
Db     420 AGTGAAGCTGTCCTGCAAGGCTTCTGGCTACACATTCACCAGCTACTGGATGCACTGGGT 479

Qy     472 GAAGCAGAGCCATGGAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGA 531
           ||||||||| |||||| |   ||||||||||||||||  |||| |||||   | ||||
Db     480 GAAGCAGAGGCATGGACAAGGCCTTGAGTGGATTGGAAATATTTATCCTGGTAGTGGTAG 539

Qy     532 TACTAACTACAACCAGAAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAG 591
           |||||||||| |  ||||||||||| ||||||||||| | ||||||||||  || |||||
Db     540 TACTAACTACGATGAGAAGTTCAAGAGCAAGGCCACACTGACTGTAGACACATCCTCCAG 599

Qy     592 CACAGCCTACATGGAGCTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGC 651
           ||||||||||||| | |||  ||| |||||||||||||||||||| |||||||||||| |
Db     600 CACAGCCTACATGCACCTCAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTAC 659

Qy     652 AAGGAGGGTTACGGACTGGTACTTCGATG---------TCTGGGGCGCAGGGACCACGGT 702
           ||| ||      | |||||  || | |||          ||||||   ||| ||| | ||
Db     660 AAGAAGCAGCCGGAACTGGGTCTACTATGCTATGGACTACTGGGGTCAAGGAACCTCAGT 719

Qy     703 CACCGTCTCC 712
           ||||||||||
Db     720 CACCGTCTCC 729


RESULT   6
```

```
I32407
LOCUS       I32407        720 bp    DNA             PAT       07-JAN-1997
DEFINITION  Sequence 2 from patent US 5587458.
ACCESSION   I32407
NID         g1823198
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 720)
  AUTHORS   King,C.Richter, Kasprzyk,P.G. and Bird,R.E.
  TITLE     Anti-erbB-2 antibodies, combinations thereof, and therapeutic and
            diagnostic uses thereof
  JOURNAL   Patent: US 5587458-A 2 24-DEC-1996;
FEATURES             Location/Qualifiers
     source          1. .720
                     /organism="unknown"
BASE COUNT      184 a    178 c    179 g    179 t
ORIGIN


  Query Match             59.3%;  Score 422;  DB 6;  Length 720;
  Best Local Similarity   79.3%;  Pred. No. 1.1e-119;
  Matches  511;  Conservative    1;  Mismatches  131;  Indels    1;  Gaps    1;

Qy      8 TGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTACACAATG 67
          |||||||||||||||||||||||||| ||||||||||||||||||| ||||| ||| |||
Db      2 TGCAGCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGT-CACCATG 60

Qy     68 ACTTGCAGGGCCACCCCAAGTGTAAGTTACATGCACTGGTATCAGCAGAAGCCAGGATCC 127
          || ||||| |||| | ||||||||||| ||||||||||||||||||||||| || |  ||
Db     61 ACCTGCAGTGCCAGCTCAAGTGTAAGTAACATGCACTGGTATCAGCAGAAGTCAAGCACC 120

Qy    128 TCCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGTCCCTGCTCGCTTC 187
          ||||||||||  ||| ||||||  ||||||||| ||:|||||||||||||| |||||||
Db    121 TCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTTC 180

Qy    188 AGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGAGGCTGAAGAT 247
          |||||| |||||||||| | |||||||||||||||  |||||||  ||||||||||||||
Db    181 AGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGAT 240

Qy    248 GCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGGAGGGGGGTCC 307
          |||||||||||||| ||  | ||| ||||| |   ||||  |||||||||   |||| |
Db    241 GCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGGACA 300

Qy    308 AAGCTGGAAATAAAAGGTTCTACCTCTGGTTCTGGTAAATCTTCTGAAGGTAAAGGTGTG 367
          ||| |||||||||||||||||||||| || ||||||||||||||||||||||||||||||
Db    301 AAGTTGGAAATAAAAGGTTCTACCTCCGGATCTGGTAAATCTTCTGAAGGTAAAGGTGTG 360

Qy    368 CAGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTCAATGAAGATATCCTGC 427
          ||||||||| |||| ||   |||| | ||       || |  | | | ||| | ||||||
Db    361 CAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGTCCTGC 420

Qy    428 AAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAAGCAGAGCCATGGA 487
          ||  ||||||  | |    | |  | | |    ||  |||||||||||||||| | || |
Db    421 AAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCCTGAA 480

Qy    488 AAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATACTAACTACAACCAG 547
           ||  ||| || |||||||||   ||| ||||| | |||||| |||||   ||  ||| |
Db    481 CAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATGACCCG 540

Qy    548 AAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCACAGCCTACATGGAG 607
          || ||| ||||||||||||   | || | |||||  || |||| | | ||||||| |  ||
Db    541 AAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTACCTTCAG 600

Qy    608 CTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGC 651
          |||  ||| ||||||||||||||| |||| ||||||||||||||
Db    601 CTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGC 644


RESULT     7
I45604
LOCUS       I45604        916 bp    DNA             PAT       29-JUL-1997
DEFINITION  Sequence 29 from patent US 5637481.
ACCESSION   I45604
NID         g2469706
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 916)
  AUTHORS   Ledbetter,J.A., Gilliland,L.K., Hayden,M.S., Linsley,P.S.,
            Bajorath,J. and Fell,H.Perry.
  TITLE     Expression vectors encoding bispecific fusion proteins and methods
            of producing biologically active bispecific fusion proteins in a
            mammalian cell
  JOURNAL   Patent: US 5637481-A 29 10-JUN-1997;
FEATURES             Location/Qualifiers
     source          1. .916
                     /organism="unknown"
BASE COUNT      254 a    223 c    228 g    211 t
ORIGIN


  Query Match             50.0%;  Score 356.2;  DB 6;  Length 916;
  Best Local Similarity   71.9%;  Pred. No. 2.3e-99;
  Matches  537;  Conservative    0;  Mismatches  173;  Indels   37;  Gaps    4;

Qy      2 TGGACCTGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTAC 61
          | ||| | ||| |||| ||| || || |   ||||||||| ||||  || || |  || |
Db     74 TCGACATCCAGATGACACAGACTACATCCTCCCTGTCTGCCTCTCTGGGAGACAGAGT-C 132

Qy     62 ACAATGACTTGCAGGGCCACCC---CAAGTGTAAGTTACATGCACTGGTATCAGCAGAAG 118
          || || | ||||||||| |  |     | |   | |||  |  ||||||||||||||||
Db    133 ACCATCAGTTGCAGGGCAAGTCAGGACATTCGCAATTATTTAAACTGGTATCAGCAGAAA 192

Qy    119 CCAGGATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGTCCCT 178
          ||||      |    ||||    ||| ||   |||||| |   |    || ||||||||
Db    193 CCAGATGGAACTGTTAAACTCCTGATCTACTACACATCAAGATTACACTCAGGAGTCCCA 252

Qy    179 GCTCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGAG 238
           |  | |||||||| |||||||||| ||   ||| ||||||||  |   ||   || |
Db    253 TCAAGGTTCAGTGGCAGTGGGTCTGGAACAGATTATTCTCTCACCATTGCCAACCTGCAA 312

Qy    239 GCTGAAGATGCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGGA 298
           | ||||||  ||||||||| |  ||||| ||| | | |      ||     ||||||||
Db    313 CCAGAAGATATTGCCACTTACTTTTGCCAACAGGGTAATACGCTTCCGTGGACGTTCGGT 372

Qy    299 GGGGGGTCCAAGCTGGAAATAAAAGGTTCTACCTCTGGTTCTGGTAAATCTTCTGAAGGT 358
          || ||  |||| |||| ||  ||| |      |  |||    |||         ||  ||
Db    373 GGAGGCACCAAACTGGTAACCAAACGGGAGCTCGGTGGCGGTGGCTCGGGCGGTGGTGGG 432

Qy    359 AAAGGT---------------------GTGCAGCTGCAGGAGTCAGGACCTGAGGTGGTG 397
             |||                     || ||||||||  |||| ||||||||  |||||
Db    433 TCGGGTGGCGGCGGATCTATCGATGAGGTCCAGCTGCAACAGTCTGGACCTGAACTGGTG 492

Qy    398 AAGCCTGGAGGTTCAATGAAGATATCCTGCAAGACTTCTGGTTACTCATTCACTGGCCAC 457
          |||||||||| ||||||||  || ||||||||| | ||||||||||||||||||||| ||
Db    493 AAGCCTGGAGCTTCAATGACAATGTCCTGCAAGGCCTCTGGTTACTCATTCACTGGCTAC 552

Qy    458 ACCATGAACTGGGTGAAGCAGAGCCATGGAAAGAACCTTGAGTGGATTGGACTTATTAAT 517
          | | |||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db    553 ATCGTGAACTGGCTGAAGCAGAGCCATGGAAAGAACCTTGAGTGGATTGGACTTATTAAT 612

Qy    518 CCTTACAATGGTGATACTAACTACAACCAGAAGTTCAAGGGCAAGGCCACATTTACTGTA 577
          || ||||| |||  ||||| |||||||||||| |||||||||||||||||||| ||||||
Db    613 CCATACAAAGGTCTTACTACCTACAACCAGAAATTCAAGGGCAAGGCCACATTAACTGTA 672

Qy    578 GACAAGTCGTCCAGCACAGCCTACATGGAGCTCCTCAGTCTGACATCTGAGGACTCTGCA 637
          |||||||| ||||||||||||||||||||||||||||||||||||||||| |||||||||
Db    673 GACAAGTCATCCAGCACAGCCTACATGGAGCTCCTCAGTCTGACATCTGAAGACTCTGCA 732

Qy    638 GTCTATTACTGTGCAAG------------GAGGGTTACGGACTGGTACTTCGATGTCTGG 685
          |||||||||||||||||              | |   |||||||||||||||||||||||
Db    733 GTCTATTACTGTGCAAGATCTGGGTACTATGGTGACTCGGACTGGTACTTCGATGTCTGG 792
```

```
Qy    686 GGCGCAGGGACCACGGTCACCGTCTCC 712
          |||||||||||||||  ||||||||||
Db    793 GGCGCAGGGACCACGTGCACCGTCTCC 819


RESULT   8
I31036
LOCUS       I31036        729 bp    DNA             PAT       20-DEC-1996
DEFINITION  Sequence 3 from patent US 5582826.
ACCESSION   I31036
NID         g1821827
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 729)
  AUTHORS   Shimamura,T., Hamuro,J., Nakazawa,H., Kanayama,Y., Sugamura,K. and
            Takeshita,T.
  TITLE     Monoclonal antibodies which bind the gamma chain of human
            interleukin-2 receptor
  JOURNAL   Patent: US 5582826-A 3 10-DEC-1996;
FEATURES             Location/Qualifiers
     source          1. .729
                     /organism="unknown"
BASE COUNT      199 a    169 c    182 g    179 t
ORIGIN


  Query Match             48.9%;  Score 348.4;  DB 6;  Length 729;
  Best Local Similarity   71.5%;  Pred. No. 5.7e-97;
  Matches  519;  Conservative    0;  Mismatches  191;  Indels   16;  Gaps    4;

Qy      1 ATGGACCTGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTA 60
          |||||  | | |||||| |||||||||||   ||| |||||||||   || || |  ||
Db      1 ATGGATATTCTGCTGACACAGTCTCCAGCCTCCCTATCTGCATCTGTGGGAGAAACTGT- 60

Qy     61 CACAATGACTTGCAGGGCCACCCCAAGTGT---AAGTTACATGCACTGGTATCAGCAGAA 117
          ||| || || ||  | || |     | | |    | |||  |    ||||||||||||||
Db     60 CACCATCACATGTCGAGCAAGTGGGAATATTCACAATTATTTAGCATGGTATCAGCAGAA 119

Qy    118 GCCAGGATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGTCCC 177
           |  |||   || ||  | |    | | ||||   ||   | | | ||   ||| || ||
Db    120 ACAGGGAAAATCTCCTCAGCTCCTGGTCTATAATGCAAAAACCTTAGCAGATGGTGTGCC 179

Qy    178 TGCTCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGA 237
            |  | ||||||||| |||| || || ||    || |||||||   ||| |||  || |
Db    180 ATCAAGGTTCAGTGGCAGTGGATCAGGAACACAATATTCTCTCAAGATCAACAGCCTGCA 239

Qy    238 GGCTGAAGATGCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGG 297
          | ||||||||  ||  | ||||||||| || || |   |  |||  ||    ||||||||
Db    240 GCCTGAAGATTTTGGGAGTTATTACTGTCAACATTTTTGGAGTACTCCGTGGACGTTCGG 299

Qy    298 AGGGGGGTCCAAGCTGGAAATAAAAGGTTCTA---CCTCTGGTTCTGGTAAATCTTCTGA 354
           || ||| ||||||||||  | ||||     |   |||| || |||||       ||  |
Db    300 TGGAGGGACCAAGCTGGAGCTCAAAGTCGAGAAATCCTCAGGATCTGGCTCCGAATCCAA 359

Qy    355 AGGTAAAGGTGTGCAGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTCAAT 414
          | | |     ||  | ||  ||||||| ||||||||| ||||||||||||  | |||| |
Db    360 AAGCACGCAGGTCAAACTCGAGGAGTCTGGACCTGAGCTGGTGAAGCCTGGGGCTTCAGT 419

Qy    415 GAAGATATCCTGCAAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAA 474
          |||||||||||||||| ||||||||||||||||||||||| ||  |||| ||||||||||
Db    420 GAAGATATCCTGCAAGGCTTCTGGTTACTCATTCACTGGCTACTACATGCACTGGGTGAA 479

Qy    475 GCAGAGCCATGGAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATAC 534
          ||| ||||||| ||||| ||||||||||||||||| |||||||||||||||||||| |||
Db    480 GCAAAGCCATGTAAAGAGCCTTGAGTGGATTGGACGTATTAATCCTTACAATGGTGCTAC 539

Qy    535 TAACTACAACCAGAAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCAC 594
          || |||||||||||| ||||||| ||||||||  || |||||||| ||||| ||||||||
Db    540 TAGCTACAACCAGAATTTCAAGGACAAGGCCAGCTTGACTGTAGATAAGTCCTCCAGCAC 599

Qy    595 AGCCTACATGGAGCTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAG 654
```

```
          ||||||||||||||||| ||| ||||||||||||||||||||||||||||||||||||||
Db    600 AGCCTACATGGAGCTCCACAGCCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAG 659

Qy    655 GAGG---------GTTACGGACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCAC 705
             |         | ||    ||     || | |  |||||||  ||||||   ||||||
Db    660 AGAGAATTACTACGGTAGTAGCTACGGGTTTGCTTACTGGGGCCAAGGGACTCTGGTCAC 719

Qy    706 CGTCTC 711
           |||||
Db    720 TGTCTC 725


RESULT   9
E10362
ID   E10362     standard; RNA; ROD; 729 BP.
XX
AC   E10362;
XX
NI   d1108699
XX
DT   08-OCT-1997 (Rel. 52, Created)
DT   08-OCT-1997 (Rel. 52, Last updated, Version 1)
XX
DE   cDNA encoding an monoclonal antibody against human Interleukin-2
DE   receptor gamma chain.
XX
KW   JP 1995313188-A/2.
XX
OS   Mus sp.
OC   Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Rodentia;
OC   Sciurognathi; Muridae; Murinae; Mus.
XX
RN   [1]
RP   1-729
RA   Shimamura T., Hamuro J., Nakazawa H., Kanayama Y., Sugamura K.,
RA   Takeshita T.;
RT   "IMMUNOSUPPRESSIVE AGENT";
RL   Patent number JP 1995313188-A/2, 05-DEC-1995.
RL   AJINOMOTO CO INC, SUGAMURA KAZUO,.
XX
CC   OS   Mus sp. (mouse)
CC   PN   JP 1995313188-A/2
CC   PD   05-DEC-1995
CC   PF   21-APR-1994 JP 1994082836
CC   PR   21-APR-1993 JP 93P    94491, 07-MAR-1994 JP 94P    36065
CC   PI   SHIMAMURA TOSHIAKI, HAMURO JUNJI, NAKAZAWA HARUMI,
CC   PI   KANAYAMA YUKA,
CC   PI   SUGAMURA KAZUO, TAKESHITA TOSHIICHI
CC   PC   C12P21/08,A61K39/395,A61K39/395,C12N1/21,C12N5/20,
CC   PC   C12N15/13//C12N15/06,
CC   PC   (C12P21/08,C12R1:19),(C12P21/08,C12R1:91),(C12N1/21,C12R1:19);
CC   FH   Key             Location/Qualifiers
CC   FH
CC   FT   source          1. .729
CC   FT                   /organism="Mus sp."
CC   FT                   /cell_type="hybridoma"
CC   FT                   /cell_line="GP-4"
CC   FT   mat_peptide     1. .729
CC   FT                   /product="anti-IL-2 receptor gamma chain"
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .729
FT                   /organism="Mus sp."
XX
SQ   Sequence 729 BP; 199 A; 169 C; 182 G; 179 T; 0 other;


  Query Match             48.9%;  Score 348.4;  DB 27;  Length 729;
  Best Local Similarity   71.5%;  Pred. No. 5.7e-97;
  Matches  519;  Conservative    0;  Mismatches  191;  Indels   16;  Gaps    4;

Qy      1 ATGGACCTGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTA 60
```

```
           |||||  | | |||||| |||||||||||   ||| |||||||||   || || |  ||
Db       1 ATGGATATTCTGCTGACACAGTCTCCAGCCTCCCTATCTGCATCTGTGGGAGAAACTGT- 60

Qy      61 CACAATGACTTGCAGGGCCACCCCAAGTGT---AAGTTACATGCACTGGTATCAGCAGAA 117
           ||| || || ||  | || |     | | |    | |||  |    ||||||||||||||
Db      60 CACCATCACATGTCGAGCAAGTGGGAATATTCACAATTATTTAGCATGGTATCAGCAGAA 119

Qy     118 GCCAGGATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGTCCC 177
            |  |||   || ||  | |    | | ||||   ||   | | | ||   ||| || ||
Db     120 ACAGGGAAAATCTCCTCAGCTCCTGGTCTATAATGCAAAAACCTTAGCAGATGGTGTGCC 179

Qy     178 TGCTCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGA 237
             |  | ||||||||| |||| || || ||    || |||||||   ||| |||  || |
Db     180 ATCAAGGTTCAGTGGCAGTGGATCAGGAACACAATATTCTCTCAAGATCAACAGCCTGCA 239

Qy     238 GGCTGAAGATGCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGG 297
           | ||||||||  ||  | ||||||||| || || |   |  |||  ||    ||||||||
Db     240 GCCTGAAGATTTTGGGAGTTATTACTGTCAACATTTTTGGAGTACTCCGTGGACGTTCGG 299

Qy     298 AGGGGGGTCCAAGCTGGAAATAAAAGGTTCTA---CCTCTGGTTCTGGTAAATCTTCTGA 354
            || ||| ||||||||||  | ||||     |   |||| || |||||       ||  |
Db     300 TGGAGGGACCAAGCTGGAGCTCAAAGTCGAGAAATCCTCAGGATCTGGCTCCGAATCCAA 359

Qy     355 AGGTAAAGGTGTGCAGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTCAAT 414
           | | |     ||  | ||  ||||||| ||||||||| ||||||||||||| | |||| |
Db     360 AAGCACGCAGGTCAAACTCGAGGAGTCTGGACCTGAGCTGGTGAAGCCTGGGGCTTCAGT 419

Qy     415 GAAGATATCCTGCAAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAA 474
           |||||||||||||||| ||||||||||||||||||||||| ||  |||| ||||||||||
Db     420 GAAGATATCCTGCAAGGCTTCTGGTTACTCATTCACTGGCTACTACATGCACTGGGTGAA 479

Qy     475 GCAGAGCCATGGAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATAC 534
           ||| ||||||| ||||| ||||||||||||||||| |||||||||||||||||||| |||
Db     480 GCAAAGCCATGTAAAGAGCCTTGAGTGGATTGGACGTATTAATCCTTACAATGGTGCTAC 539

Qy     535 TAACTACAACCAGAAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCAC 594
           || |||||||||||| ||||||| ||||||||  || |||||||| ||||| ||||||||
Db     540 TAGCTACAACCAGAATTTCAAGGACAAGGCCAGCTTGACTGTAGATAAGTCCTCCAGCAC 599

Qy     595 AGCCTACATGGAGCTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAG 654
           ||||||||||||||||| ||| ||||||||||||||||||||||||||||||||||||||
Db     600 AGCCTACATGGAGCTCCACAGCCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAG 659

Qy     655 GAGG---------GTTACGGACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCAC 705
              |         | ||    ||     || | |  |||||||  ||||||   ||||||
Db     660 AGAGAATTACTACGGTAGTAGCTACGGGTTTGCTTACTGGGGCCAAGGGACTCTGGTCAC 719

Qy     706 CGTCTC 711
            |||||
Db     720 TGTCTC 725


RESULT  10
SYNDA44X
LOCUS       SYNDA44X      744 bp    mRNA            SYN       01-MAR-1996
DEFINITION  Synthetic single-chain Fv fusion protein (DA4.4/212) mRNA,
            constructed from variable light and heavy chain regions of mouse
            monoclonal antibody DA4.4, 3' end of cds.
ACCESSION   L17037
NID         g310739
KEYWORDS    IgM-binding protein; fusion protein; fusion protein DA4.4/212;
            monoclonal antibody.
SOURCE      Artificial gene cDNA to mRNA.
  ORGANISM  artificial sequence
            artificial sequence.
REFERENCE   1  (bases 1 to 744)
  AUTHORS   Lee,T.K., Rollence,M.L., Hallberg,P.L., Oelkuct,M.S., Dodd,S.W.,
            Nagle,J.W. and Filpula,D.R.
  TITLE     Production of engineered IgM-binding single-chain antibodies in
            Escherichia coli
  JOURNAL   J. Ind. Microbiol. 14 (5), 371-376 (1995)
  MEDLINE   95336669
FEATURES             Location/Qualifiers
```

```
     source          1..744
                     /organism="artificial sequence"
                     /db_xref="taxon:29278"
     CDS             <1..735
                     /standard_name="DA4.4/212 sFv protein"
                     /codon_start=1
                     /transl_table=11
                     /function="binds human IgM"
                     /product="fusion protein"
                     /db_xref="PID:g310740"
                     /translation="DVVMTQSPSSLAMSVGQKVTMSCKSSQSLLNSSNQKNYLAWYQQ
                     KPGQSPELLVYFASTRESGVPDRFIGSGSGTDFTLTISSVQAEDLADYFCQQHYSTPF
                     TFGSGTKLEIKGSTSGSGKSSEGKGQVQLQQPGAEFVKPGAPVKLSCKASGYPFTTYW
                     VNWMKQRPGRGLEWIGRIDPYDSETLYNQKFKDKATLTVDKSSSTAYIQLSSLTSEDS
                     AVYYCARETYDYPFAYWGQGTLVTVS"
     misc_feature    1..339
                     /note="variable light chain derived from mouse monoclonal
                     antibody DA4.4 (ATCC HB57)"
     misc_feature    340..381
                     /note="linker segment"
     misc_feature    382..744
                     /note="variable heavy chain derived from mouse monoclonal
                     antibody DA4.4 (ATCC HB57)"
BASE COUNT      192 a    187 c    186 g    179 t
ORIGIN


  Query Match            47.5%;  Score 338;  DB 15;  Length 744;
  Best Local Similarity  72.6%;  Pred. No. 9.2e-94;
  Matches  450;  Conservative    1;  Mismatches  166;  Indels    3;  Gaps    1;

Qy      95 TACATGCACTGGTATCAGCAGAAGCCAGGATCCTCCCCCAAACCTTGGATTTATACCACA 154
           ||  ||  |||||| |||||||| ||||||   || ||  ||| |  | | ||     ||
Db     112 TATTTGGCCTGGTACCAGCAGAAACCAGGACAGTCTCCTGAACTTCTGGTATACTTTGCA 171

Qy     155 TCCAACCTKGCTTCTGGAGTCCCTGCTCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTAC 214
           ||||    :|  ||||| ||||||| ||||||||  ||| |||| ||||||||   || |
Db     172 TCCACTAGGGAATCTGGGGTCCCTGATCGCTTCATAGGCAGTGGATCTGGGACAGATTTC 231

Qy     215 TCTCTCACAGTCAGCAGAGTGGAGGCTGAAGATGCTGCCACTTATTACTGCCAGCAGTGG 274
            |||| ||  ||||||| ||| ||||||||||    ||   ||| | ||| |||||
Db     232 ACTCTTACCATCAGCAGTGTGCAGGCTGAAGACCTGGCAGATTACTTCTGTCAGCAACAT 291

Qy     275 AGTCGTAGCCCACCCACGTTCGGAGGGGGGTCCAAGCTGGAAATAAAAGGTTCTACCTCT 334
             | | |  |||  |||||||||   |||| | ||||| ||||||||||| ||||| ||
Db     292 TATAGCACTCCATTCACGTTCGGCTCGGGGACAAAGCTTGAAATAAAAGGCTCTACTTCC 351

Qy     335 GGTTCTGGTAAATCTTCTGAAGGTAAAGGT---GTGCAGCTGCAGGAGTCAGGACCTGAG 391
           |||   || ||||| |||||||| ||||||   || ||||||||| || | ||  |||||
Db     352 GGTAGCGGCAAATCCTCTGAAGGCAAAGGTCAGGTTCAGCTGCAGCAGCCTGGGGCTGAG 411

Qy     392 GTGGTGAAGCCTGGAGGTTCAATGAAGATATCCTGCAAGACTTCTGGTTACTCATTCACT 451
            | ||||||||||| | | || ||||  | ||||||||| ||||||| ||| | |||||
Db     412 TTTGTGAAGCCTGGGGCTCCAGTGAAACTGTCCTGCAAGGCTTCTGGCTACCCCTTCACC 471

Qy     452 GGCCACACCATGAACTGGGTGAAGCAGAGCCATGGAAAGAACCTTGAGTGGATTGGACTT 511
             | ||    |||||||| |||||||| | | ||||     ||| ||||||||||||
Db     472 ACCTACTGGGTGAACTGGATGAAGCAGCGGCCTGGACGAGGCCTCGAGTGGATTGGAAGG 531

Qy     512 ATTAATCCTTACAATGGTGATACTAACTACAACCAGAAGTTCAAGGGCAAGGCCACATTT 571
           ||| |||||||| || |||| |||  |||||| || |||||||||| |||||||||| |
Db     532 ATTGATCCTTACGATAGTGAAACTCTCTACAATCAAAAGTTCAAGGACAAGGCCACACTG 591

Qy     572 ACTGTAGACAAGTCGTCCAGCACAGCCTACATGGAGCTCCTCAGTCTGACATCTGAGGAC 631
           ||||| ||||| || |||||||||||||||||  | |||  ||| |||||||||||||||
Db     592 ACTGTGGACAAATCCTCCAGCACAGCCTACATCCAACTCAGCAGCCTGACATCTGAGGAC 651

Qy     632 TCTGCAGTCTATTACTGTGCAAGGAGGGTTACGGACTGGTACTTCGATGTCTGGGGCGCA 691
           ||||| ||||||||||||||||||     |   || |    ||| | |  |||||||  |
Db     652 TCTGCGGTCTATTACTGTGCAAGGGAAACTTATGATTACCCCTTTGCTTACTGGGGCCAA 711

Qy     692 GGGACCACGGTCACCGTCTC 711
           |||||   |||||| || ||
Db    712 GGGACTCTGGTCACTGTGTC 731


RESULT  11
SYNAR3IGA
LOCUS       SYNAR3IGA     738 bp    DNA             SYN       15-SEP-1990
DEFINITION  Synthetic mouse Ig heavy and light chain antibody (OVB3), partial
            cds.
ACCESSION   M34000
NID         g207974
KEYWORDS    C-region.
SOURCE      Mouse hybridoma, cDNA to mRNA, and E.coli vector DNA.
  ORGANISM  artificial sequence
            artificial sequence.
REFERENCE   1  (bases 1 to 738)
  AUTHORS   Chaudhary,V.K., Batra,J.K., Gallo,M.G., Willingham,M.C.,
            Fitzgerald,D.J. and Pastan,I.
  TITLE     A rapid method of cloning functional variable-region antibody genes
            in Escherichia coli as single-chain immunotoxins
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 87, 1066-1070 (1990)
  MEDLINE   90138938
FEATURES             Location/Qualifiers
     source          1..738
                     /organism="artificial sequence"
                     /db_xref="taxon:29278"
     CDS             <1..>738
                     /note="OVB3 Ig light and heavy chains"
                     /codon_start=1
                     /transl_table=11
                     /db_xref="PID:g207975"
                     /translation="HMENVLTQSPAIMSASPGEKVTMTCSAISTVSYMHWYQQKSSTS
                     PKLWIYETSKLASGVPGRFSGSGSGNSYSLTISSMEAEDVATYYCFQGSGYPYTFGGG
                     TKLELKEGKSSGSGSESKVDDVQLVESGGGLVQPGGSRQLSCAASGFTFSSFGMHWVR
                     QAPEKGLEWVAYISRGGNTIYYANTVKGRFTISRDNPKNTLFLQMTSLRSDDTAMYYC
                     ARSHYYGYFYAMDYWGQGTTLTVSSKAF"
     misc_feature    326..365
                     /note="linker DNA"
BASE COUNT      181 a    193 c    192 g    172 t
ORIGIN


  Query Match             45.0%;  Score 320.6;  DB 15;  Length 738;
  Best Local Similarity   67.6%;  Pred. No. 2.1e-88;
  Matches  479;  Conservative    1;  Mismatches  225;  Indels    4;  Gaps    2;

Qy      1 ATGGACCTGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTA 60
          |||||      ||| |||||||||||||||||| ||||||||||||||||||| |||||
Db      4 ATGGAAAATGTGCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGT- 63

Qy     61 CACAATGACTTGCAGGGCCACCCCAAGTGTAAGTTACATGCACTGGTATCAGCAGAAGCC 120
          ||| ||||| ||||| |||| | ||| ||||||||||||||||||||| |||||||||.|
Db     63 CACCATGACCTGCAGTGCCATCTCAACTGTAAGTTACATGCACTGGTACCAGCAGAAGTC 122

Qy    121 AGGATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGTCCCTGC 180
          | |  ||||||||||||  |||||||||   |||||||| ||:|||||||||||||| |
Db    123 AAGCACCTCCCCCAAACTCTGGATTTATGAGACATCCAAACTGGCTTCTGGAGTCCCAGG 182

Qy    181 TCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGAGGC 240
          ||||||||||||| |||||||||| | |||||||||||||||  |||||||  |||||||
Db    183 TCGCTTCAGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGC 242

Qy    241 TGAAGATGCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGGAGG 300
          |||||||| |||||||||||||||    ||| ||||| |   |||   ||||||||||||
Db    243 TGAAGATGTTGCCACTTATTACTGTTTTCAGGGGAGTGGGTACCCGTACACGTTCGGAGG 302

Qy    301 GGGGTCCAAGCTGGAAATAAAA---GGTTCTACCTCTGGTTCTGGTAAATCTTCTGAAGG 357
          |||| ||||||||||  | |||   |||    |||| || |||||       ||  |||
Db    303 GGGGACCAAGCTGGAGCTCAAAGAGGGTAAATCCTCAGGATCTGGCTCCGAATCCAAAGT 362

Qy    358 TAAAGGTGTGCAGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTCAATGAA 417
            | | ||||||||||  |||||| ||    |   | ||| |||||||||| ||   | |
Db    363 CGACGATGTGCAGCTGGTGGAGTCTGGGGGAGGCTTAGTGCAGCCTGGAGGGTCCCGGCA 422
```

```
Qy    418 GATATCCTGCAAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAAGCA 477
            | |||||     | ||||| | | | |||| | ||      ||| |||||||    ||
Db    423 ACTCTCCTGTGCAGCCTCTGGATTCACTTTCAGTAGCTTTGGAATGCACTGGGTTCGTCA 482

Qy    478 GAGCCATGGAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATACTAA 537
          |   |  |  |||   || |||||| | | |   ||||.|| |  |  |  |   |   |
Db    483 GGCTCCAGAGAAGGGGCTGGAGTGGGTCGCATACATTAGTCGTGGCGGTAATACCATCTA 542

Qy    538 CTACAACCAGAAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCACAGC 597
          |||     | |   | ||||||     |||  |  |   ||||||  |    | |||
Db    543 CTATGCAAACACAGTGAAGGGCCGATTCACCATCTCCAGAGACAATCCCAAGAACACCCT 602

Qy    598 CTACATGGAGCTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGGAG 657
           | | || |  |   |||||| |  ||||| ||| | ||  | ||||||||||||||
Db    603 GTTCCTGCAAATGACCAGTCTAAGGTCTGACGACACGGCCATGTATTACTGTGCAAGATC 662

Qy    658 GGTTACGGACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACC 706
             |    || | |||||| |||    || |    ||     ||  |||
Db    663 CCATTACTACGGCTACTTCTATGCTATGGACTACTGGGGTCAGGGAACC 711


RESULT  12
AR012822
LOCUS       AR012822      797 bp    DNA             PAT       04-DEC-1998
DEFINITION  Sequence 3 from patent US 5763733.
ACCESSION   AR012822
NID         g3971140
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 797)
  AUTHORS   Whitlow,M., Filpula,D. and Shorr,R.
  TITLE     Antigen-binding fusion proteins
  JOURNAL   Patent: US 5763733-A 3 09-JUN-1998;
FEATURES             Location/Qualifiers
     source          1..797
                     /organism="unknown"
BASE COUNT      196 a    203 c    195 g    203 t
ORIGIN


  Query Match             44.8%;  Score 319;  DB 6;  Length 797;
  Best Local Similarity   72.0%;  Pred. No. 6.7e-88;
  Matches  447;  Conservative    1;  Mismatches  161;  Indels   12;  Gaps    2;

Qy     95 TACATGCACTGGTATCAGCAGAAGCCAGGATCCTCCCCCAAACCTTGGATTTATACCACA 154
          ||| ||  |||||| |||||||| |||||    || || ||||    ||||||     ||
Db    112 TACTTGGCCTGGTACCAGCAGAAACCAGGGCAGTCTCCTAAACTGCTGATTTACTGGGCA 171

Qy    155 TCCAACCTKGCTTCTGGAGTCCCTGCTCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTAC 214
          |||     :|  ||||| ||||||| ||||||||  ||| |||| ||||||||   || |
Db    172 TCCGCTAGGGAATCTGGGGTCCCTGATCGCTTCACAGGCAGTGGATCTGGGACAGATTTC 231

Qy    215 TCTCTCACAGTCAGCAGAGTGGAGGCTGAAGATGCTGCCACTTATTACTGCCAGCAGTGG 274
           ||||| |  ||||||| ||| || |||||||    ||   ||||||||| |||||||
Db    232 ACTCTCTCCATCAGCAGTGTGAAGACTGAAGACCTGGCAGTTTATTACTGTCAGCAGTAT 291

Qy    275 AGTCGTAGCCCACCCACGTTCGGAGGGGGGTCCAAGCTGGAAATAAAAGGTTCTACCTCT 334
            | |    || | ||||||||| |  ||| ||||||| |   | ||||| ||||| ||
Db    292 TATAGCTATCCCCTCACGTTCGGTGCTGGGACCAAGCTTGTGCTGAAAGGCTCTACTTCC 351

Qy    335 GGTTCTGGTAAATCTTCTGAAGGTAAAGGT---GTGCAGCTGCAGGAGTCAGGACCTGAG 391
          |||   || ||||| |||||||| ||||||   || ||||||||| |||| |   |||||
Db    352 GGTAGCGGCAAATCCTCTGAAGGCAAAGGTCAGGTTCAGCTGCAGCAGTCTGACGCTGAG 411

Qy    392 GTGGTGAAGCCTGGAGGTTCAATGAAGATATCCTGCAAGACTTCTGGTTACTCATTCACT 451
           ||||||| ||||| | |||| ||||||| ||||||||| ||||||| ||| | ||||||
Db    412 TTGGTGAAACCTGGGGCTTCAGTGAAGATTTCCTGCAAGGCTTCTGGCTACACCTTCACT 471

Qy    452 GGCCACACCATGAACTGGGTGAAGCAGAGCCATGGAAAGAACCTTGAGTGGATTGGACTT 511
          | |||  | ||  |||||||||| |||| || || | ||  ||| || |||||||||  |
Db    472 GACCATGCAATTCACTGGGTGAAACAGAACCCTGAACAGGGCCTGGAATGGATTGGATAT 531
```

```
Qy    512 ATTAATCCTTACAATGGTGATACTAACTACAACCAGAAGTTCAAGGGCAAGGCCACATTT 571
           ||  |||    |||| ||||  ||| |||||  ||| ||||||||||||||||||| |
Db    532 TTTTCTCCCGGAAATGATGATTTTAAATACAATGAGAGGTTCAAGGGCAAGGCCACACTG 591

Qy    572 ACTGTAGACAAGTCGTCCAGCACAGCCTACATGGAGCTCCTCAGTCTGACATCTGAGGAC 631
          |||| |||||| || |||||||| |||||| || |||||  ||| ||||||||||||||
Db    592 ACTGCAGACAAATCCTCCAGCACTGCCTACGTGCAGCTCAACAGCCTGACATCTGAGGAT 651

Qy    632 TCTGCAGTCTATTACTGTGCAAGGAGGGTTACGGACTGGTACTTCGATGTCTGGGGCGCA 691
          |||||||| |||| |||| ||||            |  | |  | |    ||||||   |
Db    652 TCTGCAGTGTATTTCTGTACAAG---------ATCCCTGAATATGGCCTACTGGGGTCAA 702

Qy    692 GGGACCACGGTCACCGTCTCC 712
          || ||| ||||||||||||||
Db    703 GGAACCTCGGTCACCGTCTCC 723


RESULT  13
AR012824
LOCUS       AR012824       803 bp    DNA             PAT       04-DEC-1998
DEFINITION  Sequence 7 from patent US 5763733.
ACCESSION   AR012824
NID         g3971142
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 803)
  AUTHORS   Whitlow,M., Filpula,D. and Shorr,R.
  TITLE     Antigen-binding fusion proteins
  JOURNAL   Patent: US 5763733-A 7 09-JUN-1998;
FEATURES             Location/Qualifiers
     source          1. .803
                     /organism="unknown"
BASE COUNT      198 a    205 c    196 g    204 t
ORIGIN


  Query Match             44.6%;  Score 317.8;  DB 6;  Length 803;
  Best Local Similarity   73.9%;  Pred. No. 1.6e-87;
  Matches  416;  Conservative    1;  Mismatches  143;  Indels    3;  Gaps    1;

Qy     95 TACATGCACTGGTATCAGCAGAAGCCAGGATCCTCCCCCAAACCTTGGATTTATACCACA 154
          ||| ||  |||||| |||||||| |||||    || || ||||    ||||||     ||
Db    181 TACTTGGCCTGGTACCAGCAGAAACCAGGGCAGTCTCCTAAACTGCTGATTTACTGGGCA 240

Qy    155 TCCAACCTKGCTTCTGGAGTCCCTGCTCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTAC 214
          |||     :|  ||||| ||||||| ||||||||  ||| |||| ||||||||   || |
Db    241 TCCGCTAGGGAATCTGGGGTCCCTGATCGCTTCACAGGCAGTGGATCTGGGACAGATTTC 300

Qy    215 TCTCTCACAGTCAGCAGAGTGGAGGCTGAAGATGCTGCCACTTATTACTGCCAGCAGTGG 274
           ||||| |  ||||||| ||| || |||||||    ||   ||||||||| |||||||
Db    301 ACTCTCTCCATCAGCAGTGTGAAGACTGAAGACCTGGCAGTTTATTACTGTCAGCAGTAT 360

Qy    275 AGTCGTAGCCCACCCACGTTCGGAGGGGGGTCCAAGCTGGAAATAAAAGGTTCTACCTCT 334
            | |    || | ||||||||| |  ||| ||||||| |   | ||||| ||||| ||
Db    361 TATAGCTATCCCCTCACGTTCGGTGCTGGGACCAAGCTTGTGCTGAAAGGCTCTACTTCC 420

Qy    335 GGTTCTGGTAAATCTTCTGAAGGTAAAGGT---GTGCAGCTGCAGGAGTCAGGACCTGAG 391
          |||   || ||||| |||||||| ||||||   || ||||||||| |||| |   |||||
Db    421 GGTAGCGGCAAATCCTCTGAAGGCAAAGGTCAGGTTCAGCTGCAGCAGTCTGACGCTGAG 480

Qy    392 GTGGTGAAGCCTGGAGGTTCAATGAAGATATCCTGCAAGACTTCTGGTTACTCATTCACT 451
           ||||||| ||||| | |||| ||||||| ||||||||| ||||||| ||| | ||||||
Db    481 TTGGTGAAACCTGGGGCTTCAGTGAAGATTTCCTGCAAGGCTTCTGGCTACACCTTCACT 540

Qy    452 GGCCACACCATGAACTGGGTGAAGCAGAGCCATGGAAAGAACCTTGAGTGGATTGGACTT 511
          | |||  | ||  |||||||||| |||| || || | ||  ||| || |||||||||  |
Db    541 GACCATGCAATTCACTGGGTGAAACAGAACCCTGAACAGGGCCTGGAATGGATTGGATAT 600

Qy    512 ATTAATCCTTACAATGGTGATACTAACTACAACCAGAAGTTCAAGGGCAAGGCCACATTT 571
           ||  |||    |||| ||||  ||| |||||  ||| ||||||||||||||||||| |
```

```
Db    601 TTTTCTCCCGGAAATGATGATTTTAAATACAATGAGAGGTTCAAGGGCAAGGCCACACTG 660

Qy    572 ACTGTAGACAAGTCGTCCAGCACAGCCTACATGGAGCTCCTCAGTCTGACATCTGAGGAC 631
          |||| |||||| || |||||||| |||||| || |||||  ||| ||||||||||||||
Db    661 ACTGCAGACAAATCCTCCAGCACTGCCTACGTGCAGCTCAACAGCCTGACATCTGAGGAT 720

Qy    632 TCTGCAGTCTATTACTGTGCAAG 654
          |||||||| |||| |||| ||||
Db    721 TCTGCAGTGTATTTCTGTACAAG 743


RESULT  14
E13598
LOCUS       E13598         735 bp    DNA             PAT       27-APR-1998
DEFINITION  DNA encoding a single chain Fv antibody of anti-T3 antibody TT1.
ACCESSION   E13598
NID         g3252400
KEYWORDS    JP 1997220092-A/1.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 735)
  AUTHORS   Ekida,T., Yasukawa,K., Imanaka,T. and Takagi,M. .
  TITLE     PRODUCTION OF SINGLE-STRAND FV ANTIBODY
  JOURNAL   Patent: JP 1997220092-A 1 26-AUG-1997;
            TOSOH CORP
COMMENT     OS   None
            OC   Artificial sequences.
            PN   JP 1997220092-A/1
            PD   26-AUG-1997
            PF   15-FEB-1996 JP 1996027622
            PI   EKIDA TEIJI, YASUKAWA KIYOSHI, IMANAKA TADAYUKI, PI   TAKAGI
            MASAHIRO
            PC   C12N15/09,C12N1/21,C12P21/08,(C12N1/21,C12R1:19),(C12P21/08,
            PC   C12R1:19);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .735
            FT                   /organism='Artificial sequences'.
FEATURES             Location/Qualifiers
     source          1. .735
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      165 a    192 c    197 g    181 t
ORIGIN


  Query Match             44.4%;  Score 315.8;  DB 6;  Length 735;
  Best Local Similarity   68.6%;  Pred. No. 6.5e-87;
  Matches  499;  Conservative    1;  Mismatches  208;  Indels   19;  Gaps    4;

Qy      4 GACCTGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTACAC 63
          ||| | |||||||||||||||||| | | | ||  ||||||||| ||||||||| | |||
Db      7 GACATTCAGCTGACCCAGTCTCCAACCACCATGGTTGCATCTCCCGGGGAGAAGAT-CAC 65

Qy     64 AATGACTTGCAGGGCCACCCCAAGTGTAAG------TTACATGCACTGGTATCAGCAGAA 117
           || || ||||| |||| | ||||| ||||      |||| |||| ||||||||||||||
Db     66 TATCACCTGCAGTGCCAGCTCAAGTATAAGTTCCAATTACTTGCATTGGTATCAGCAGAA 125

Qy    118 GCCAGGATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGTCCC 177
          |||||||| |||||| ||||  | ||||||||  |||||||| ||:|||||||| |||||
Db    126 GCCAGGATTCTCCCCTAAACTCTTGATTTATAGGACATCCAATCTGGCTTCTGGTGTCCC 185

Qy    178 TGCTCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGA 237
            |||||||||||||| ||||||||||||||||||||||||||||| |  |||   ||||
Db    186 AACTCGCTTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATTGGCACCATGGA 245

Qy    238 GGCTGAAGATGCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGG 297
          ||||||||||| ||||||||| |||||||||||| | ||| |||  || | |||||||||
Db     246 GGCTGAAGATGTTGCCACTTACTACTGCCAGCAGGGTAGTAGTATACCGCTCACGTTCGG 305

Qy     298 AGGGGGGTCCAAGCTGGAAATAAAAGGTTCTACCTCTGGTTCTGGTAAATCTTCTGAAGG 357
            |  ||| ||||||| || ||    ||   |  ||| ||   ||||   ||   ||  ||
Db     306 TGCTGGGACCAAGCTCGAGATCGGTGGCGGTGGCTCGGGCGGTGGTGGGTCGGGTGGCGG 365

Qy     358 TAAA---GGTGTGCAGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTCAAT 414
              |   |  ||  ||||||||||||| ||    |   | ||||||| ||| || ||  |
Db     366 CGGATCCGAGGTCAAGCTGCAGGAGTCTGGGGGAGGCTTAGTGAAGCTTGGCGGGTCCCT 425

Qy     415 GAAGATATCCTGCAAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAA 474
           |||  | |||||  |  | ||||| | | | |||| | |  |   ||||   |||||
Db     426 GAAACTCTCCTGTGAAGCCTCTGGATTCACTTTCAGTAGTTATTACATGTCTTGGGTTCG 485

Qy     475 GCAGAGCCATGGAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATAC 534
            ||||  |  |  ||||  || |||| | | | |   ||||||  | |   ||||   ||
Db     486 CCAGACTCCAGAGAAGAGGCTGGAGTTGGTCGCAGCCATTAATAGTAATGGTGGTACCAC 545

Qy     535 TAACTACAACCAGAAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCAC 594
             ||||     | |   | ||||||     |||  |  |   ||||||  |    | |||
Db     546 CTACTATTCAGACACTGTGAAGGGCCGATTCACCATCTCCAGAGACAATGCCAAGAACAC 605

Qy     595 AGCCTACATGGAGCTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAG 654
               ||| || |  |   ||||||||  ||||||||| | ||  | ||||||||||||||
Db     606 CCTGTACCTGCAAATGAGCAGTCTGAAGTCTGAGGACACAGCCTTGTATTACTGTGCAAG 665

Qy     655 GAGGGTTACGGACTGGTACTTCGATGT---------CTGGGGCGCAGGGACCACGGTCAC 705
              |||  |  | |  ||| || ||||         |||||||  |||||||||||||||
Db     666 CCCGGTCTCCTATTATTACCTCTATGTCATGTACTACTGGGGCCAAGGGACCACGGTCAC 725

Qy     706 CGTCTCC 712
           |||||||
Db     726 CGTCTCC 732


RESULT  15
I08678
LOCUS       I08678        741 bp                   PAT        14-NOV-1994
DEFINITION  Sequence 10 from Patent WO 8801649.
ACCESSION   I08678
NID         g588616
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 741)
  AUTHORS   Ladner,R.C. and Bird,R.E.
  TITLE     SINGLE POLYPEPTIDE CHAIN BINDING MOLECULES
  JOURNAL   Patent: WO 8801649-A 10 10-MAR-1988;
FEATURES             Location/Qualifiers
     source          1. .741
                     /organism="unknown"
BASE COUNT      172 a    188 c    192 g    189 t
ORIGIN


  Query Match             44.0%;  Score 313.2;  DB 6;  Length 741;
  Best Local Similarity   69.8%;  Pred. No. 4.1e-86;
  Matches  469;  Conservative    0;  Mismatches  193;  Indels   10;  Gaps    3;

Qy       1 ATGGACCTGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTA 60
           |||||      ||| |||||||||| ||||||| ||||||||||||||||||| |||||
Db       1 ATGGAAAATGTGCTCACCCAGTCTCGAGCAATCATGTCTGCATCTCCAGGGGAAAAGGT- 60

Qy      61 CACAATGACTTGCAGGGCCACCCCAAGTGTA------AGTTACATGCACTGGTATCAGCA 114
           ||| ||||| |||||||||| | ||||||||      |||||| |||||||||  |||||
Db      60 CACCATGACCTGCAGGGCCAGCTCAAGTGTAAGTTCCAGTTACTTGCACTGGTTCCAGCA 119

Qy     115 GAAGCCAGGATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGT 174
           |||| ||||  ||||||||||||  |||||||||| |||||||||| | |||||||||||
Db     120 GAAGTCAGGTGCCTCCCCCAAACTCTGGATTTATAGCACATCCAACTTCGCTTCTGGAGT 179

Qy     175 CCCTGCTCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGT 234
           ||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||| ||
Db     180 CCCTGCTCGCTTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGTGT 239

Qy     235 GGAGGCTGAAGATGCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTT 294
           |||||||||||||||||||||||||||||||||| |  |  || ||  ||||| ||||||
Db     240 GGAGGCTGAAGATGCTGCCACTTATTACTGCCAGTACAGTGGTGGTTACCCACTCACGTT 299

Qy     295 CGGAGGGGGGTCCAAGCTGGAAATAAAAGGTTCTACCTCTGGTTCTGGTAAATCT---TC 351
           ||| |  ||| |||||||| ||  |   ||||||   |||  |          ||   |
Db     300 CGGTGCTGGGACCAAGCTGAAAGAATCTGGTTCTGTTTCTTCTGAACAGCTGGCTCAGTT 359

Qy     352 TGAAGGTAAAGGTGTGCAGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTC 411
           |     |   | ||||||||||  |||||| ||    ||  | |||||||||||||| ||
Db     360 TCGTTCTCTGGATGTGCAGCTGGTGGAGTCTGGGGGAGACTTAGTGAAGCCTGGAGGGTC 419

Qy     412 AATGAAGATATCCTGCAAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGT 471
             ||||  | |||||     | ||||| | | | |||| | || |   ||||   |||||
Db     420 CCTGAAACTCTCCTGTGCAGCCTCTGGATTCACTTTCATTAGCTATGGCATGTCTTGGGT 479

Qy     472 GAAGCAGAGCCATGGAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGA 531
               |  |  |  |  ||||  || |||||| | | |   |||| |  |   | |  | |
Db     480 TCGCCGAACTCCAGACAAGAGGCTGGAGTGGGTCGCAACCATTAGTAGTGGTAGTACTTA 539

Qy     532 TACTAACTACAACCAGAAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAG 591
            ||  ||||     | |   | |||||      |||  |  |   ||||||  |    |
Db     540 CACCTACTATCCAGACAGTGTGAAGGGGCGATTCACCATCTCCAGAGACAATGCCAAGAA 599

Qy     592 CACAGCCTACATGGAGCTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGC 651
           |||    ||| || |  |   | ||||||  ||||||||| | ||  | |||||||||||
Db     600 CACCCTGTACCTGCAAATGAGCGGTCTGAAGTCTGAGGACACAGCCATGTATTACTGTGC 659

Qy     652 AAGGAGGGTTAC 663
           |||  || ||||
Db     660 AAGACGGATTAC 671


Search completed: May 13, 1999, 09:54:13
Job time: 1170 sec
```

```
                    GenCore version 4.5
          Copyright (c) 1993 - 1998  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:          May 13, 1999, 09:34:42 ; Search time 341.74 Seconds
                                           (without alignments)
                                           3178.154 Million cell updates/sec

Title:           US-08-704-178-1
Perfect score:   712
Sequence:        1 ATGGACCTGCAGCTGACCCA..........GGACCACGGTCACCGTCTCC 712

Scoring table:   IDENTITY_NUC

Searched:        2002476 seqs, 762712212 residues

Database :       EST:*
                 1:  em_est1:*
                 2:  em_est2:*
                 3:  em_est3:*
                 4:  em_est4:*
                 5:  em_est5:*
                 6:  em_est6:*
                 7:  em_est7:*
                 8:  em_est8:*
                 9:  em_est9:*
                 10:  gb_est1:*
                 11:  gb_est2:*
                 12:  gb_est3:*
                 13:  gb_est4:*
                 14:  gb_est5:*
                 15:  gb_est6:*
                 16:  gb_est7:*
                 17:  gb_est8:*
                 18:  gb_est9:*
                 19:  gb_est10:*
                 20:  gb_est11:*
                 21:  gb_est12:*
                 22:  gb_est13:*
                 23:  gb_est22:*
                 24:  gb_est14:*
                 25:  gb_est15:*
                 26:  gb_est16:*
                 27:  gb_est17:*
                 28:  gb_est18:*
                 29:  gb_est19:*
                 30:  gb_est20:*
                 31:  gb_est21:*
                 32:  em_est10:*
                 33:  em_est11:*
                 34:  em_est12:*
                 35:  em_est13:*
                 36:  em_est14:*
                 37:  em_est15:*
                 38:  em_est16:*
                 39:  em_est17:*
                 40:  em_est18:*
                 41:  em_est19:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                    SUMMARIES
                  %
Result          Query
  No.   Score   Match Length DB  ID                            Description
 ---------------------------------------------------------------------------
    1   238.2    33.5    396  25  AA691311                     AA691311 vs14f01.r
    2   190.4    26.7    344  22  AA592800                     AA592800 vo25g11.r
    3   173.6    24.4    310  30  AI180569                     AI180569 uc70f09.r
    4   162.4    22.8    294  28  AI007196                     AI007196 ua73g04.r
c   5     162    22.8    461  31  AI201426                     AI201426 qs73c06.x
    6     146    20.5    387  22  AA569186                     AA569186 nm30d10.s
    7   128.4    18.0    345  25  AA710291                     AA710291 vt53a04.r
    8     119    16.7    431  25  AA710970                     AA710970 vt93h09.r
    9   118.6    16.7    393  11  T27593                       T27593 EST100653 H
   10     118    16.6    379  21  AA291381                     AA291381 zt44g02.r
   11   116.2    16.3    335  17  AA318377                     AA318377 EST20620
   12   114.2    16.0    413  17  AA301347                     AA301347 EST14279
   13   112.4    15.8    336  29  AA987559                     AA987559 or83g01.s
   14   110.8    15.6    403  18  AA345486                     AA345486 EST51505
   15   110.8    15.6    398  24  AA423447                     AA423447 ve80a03.r
   16   110.2    15.5    488  11  H25625                       H25625 yl48g05.r1
   17   110.2    15.5    400  20  AA456778                     AA456778 zw27f11.r
   18   109.4    15.4    253  17  AA295941                     AA295941 EST101165
   19   109.4    15.4    333  17  AA300732                     AA300732 EST13847
   20   107.6    15.1    370  17  AA295093                     AA295093 EST100400
   21   106.4    14.9    363  18  AA367405                     AA367405 EST78511
   22   106.2    14.9    470  12  H62115                       H62115 yu40h01.r1
   23     106    14.9    442  24  AA405415                     AA405415 zu56d02.r
   24   105.8    14.9    328  17  AA295311                     AA295311 EST100471
   25   105.4    14.8    279  17  AA300571                     AA300571 EST13661
   26   105.4    14.8    345  18  AA335086                     AA335086 EST39457
   27   105.2    14.8    823  16  AA170256                     AA170256 ms87g10.r
c  28   103.6    14.6    209  11  T28938                       T28938 EST61186 Ho
   29   102.8    14.4    349  18  AA379044                     AA379044 EST91999
   30   102.8    14.4    418  21  AA515239                     AA515239 ng69c07.s
c  31   102.2    14.4    471  31  AI233978                     AI233978 EST230666
   32   101.8    14.3    372  25  AA581192                     AA581192 nd38b11.r
   33   101.4    14.2    382  17  AA295786                     AA295786 EST100987
   34     101    14.2    463  11  R69532                       R69532 yj82d09.r1
   35   100.8    14.2    287  21  AA496610                     AA496610 zv38g06.r
   36   100.4    14.1    210  11  R69482                       R69482 yj83c03.r1
   37   100.4    14.1    352  17  AA301261                     AA301261 EST14181
   38     100    14.0    420  30  AI120005                     AI120005 uc25e04.r
   39    99.8    14.0    432  31  AI268604                     AI268604 qo47a10.x
   40    99.2    13.9    363  20  AA464313                     AA464313 zx78c12.r
   41      99    13.9    242  21  AA479857                     AA479857 zu35b05.r
   42    96.8    13.6    339  25  AA710249                     AA710249 vt49e01.r
   43    96.2    13.5    503  10  R28232                       R28232 yh57e08.r1
   44      96    13.5    291  18  AA377074                     AA377074 EST89603
   45    95.4    13.4    287  11  T27609                       T27609 EST101034 H

                                   ALIGNMENTS

RESULT   1
AA691311
LOCUS       AA691311      396 bp    mRNA             EST       16-DEC-1997
DEFINITION  vs14f01.r1 Barstead mouse irradiated colon MPLRB7 Mus musculus cDNA
            clone 1138201 5' similar to gb:X67211 M.musculus rearranged
            immunoglobulin kappa light chain (MOUSE);, mRNA sequence.
ACCESSION   AA691311
NID         g2692247
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 396)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:619473
            Seq primer: -28m13 rev2 ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .396
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; Tissue obtained
                     from 8 week old mouse.  Colon was harvested 72 hours after
                     irradiation with 1400 Gys.  1st strand cDNA was primed
                     with  a Not I - oligo(dT) primer
                     [5'TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTT
                     T 3'];  double-stranded cDNA was ligated to Eco RI
                     adaptors [AATTCGGATCCTTG],  digested with Not I and cloned
                     into the Not I and Eco RI sites of  the modified pT7T3
                     vector. Library constructed by Bob Barstead."
                     /db_xref="taxon:10090"
                     /clone="1138201"
                     /clone_lib="Barstead mouse irradiated colon MPLRB7"
                     /dev_stage="8 weeks"
                     /lab_host="DH10B"
BASE COUNT      104 a    107 c     91 g     94 t
ORIGIN


  Query Match              33.5%;  Score 238.2;  DB 25;  Length 396;
  Best Local Similarity    89.6%;  Pred. No. 1e-64;
  Matches  266;  Conservative    1;  Mismatches   29;  Indels    1;  Gaps    1;

Qy       13 CTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTACACAATGACTTG 72
            || |||||||||||||||||| ||||||||||||||||||||||||| ||| ||||| ||
Db      101 CTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAGAAGGT-CACCATGACCTG 159

Qy       73 CAGGGCCACCCCAAGTGTAAGTTACATGCACTGGTATCAGCAGAAGCCAGGATCCTCCCC 132
            ||| |||| | ||||||||||||||||||||||||| ||||||||| ||||  |||||||
Db      160 CAGTGCCAGCTCAAGTGTAAGTTACATGCACTGGTACCAGCAGAAGTCAGGCACCTCCCC 219

Qy      133 CAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGTCCCTGCTCGCTTCAGTGG 192
            ||||   |||||||||  ||||||||| ||:|||||||||||||||||||||||||||||
Db      220 CAAAAGATGGATTTATGACACATCCAAACTGGCTTCTGGAGTCCCTGCTCGCTTCAGTGG 279

Qy      193 CGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGAGGCTGAAGATGCTGC 252
            | ||||||||||||||||||||||||||||| |||||||  |||||||||||||||||||
Db      280 CAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGCATGGAGGCTGAAGATGCTGC 339

Qy      253 CACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGGAGGGGGGTCCAA 309
            ||||||||||||||||||||||||| ||| ||| | ||||||||| |  ||| ||||
Db      340 CACTTATTACTGCCAGCAGTGGAGTAGTAACCCGCTCACGTTCGGTGCTGGGACCAA 396


RESULT    2
AA592800
LOCUS       AA592800         344 bp    mRNA             EST       17-SEP-1997
DEFINITION  vo25g11.r1 Barstead mouse irradiated colon MPLRB7 Mus musculus cDNA
            clone 1050980 5' similar to gb:X67211 M.musculus rearranged
            immunoglobulin kappa light chain (MOUSE);, mRNA sequence.
ACCESSION   AA592800
NID         g2406463
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 344)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:582556
            Trace considered overall poor quality
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .344
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; Tissue obtained
                     from 8 week old mouse.  Colon was harvested 72 hours after
                     irradiation with 1400 Gys.  1st strand cDNA was primed
                     with  a Not I - oligo(dT) primer
                     [5'TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTT
                     T 3'];  double-stranded cDNA was ligated to Eco RI
                     adaptors [AATTCGGATCCTTG],  digested with Not I and cloned
                     into the Not I and Eco RI sites of  the modified pT7T3
                     vector. Library constructed by Bob Barstead."
                     /db_xref="taxon:10090"
                     /clone="1050980"
                     /clone_lib="Barstead mouse irradiated colon MPLRB7"
                     /dev_stage="8 weeks"
                     /lab_host="DH10B"
     mRNA            <1. .>344
BASE COUNT       91 a     91 c     70 g     92 t
ORIGIN


  Query Match              26.7%;  Score 190.4;  DB 22;  Length 344;
  Best Local Similarity    83.7%;  Pred. No. 1.1e-49;
  Matches  226;  Conservative    1;  Mismatches   42;  Indels    1;  Gaps    1;

Qy       13 CTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAG-GGGAGAAGGTACACAATGACTT 71
            || || || |||||||||||| ||||||| ||||||| ||||| |    ||| | ||| |
Db       75 CTCACTCAATCTCCAGCAATCATGTCTGCTTCTCCAGAGGGAGCAATGTCACCACGACCT 134

Qy       72 GCAGGGCCACCCCAAGTGTAAGTTACATGCACTGGTATCAGCAGAAGCCAGGATCCTCCC 131
            |||| ||||   | | ||||||||||||||| ||||| ||||||||| ||||  ||||||
Db      135 GCAGTGCCAGTTCTACTGTAAGTTACATGCAATGGTAACAGCAGAAGTCAGGTACCTCCC 194

Qy      132 CCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGTCCCTGCTCGCTTCAGTG 191
            ||||| | |||||||||   ||||| || ||:||||||||| ||||||||||||||||||
Db      195 CCAAAACATGGATTTATGATACATCTAAACTGGCTTCTGGAATCCCTGCTCGCTTCAGTG 254

Qy      192 GCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGAGGCTGAAGATGCTG 251
            || || | ||||||||||| |||||||||||| |||||||| | ||||||||| ||||||
Db      255 GCAGTAGCTCTGGGACCTCCTACTCTCTCACAATCAGCAGAATTGAGGCTGAACATGCTG 314

Qy      252 CCACTTATTACTGCCAGCAGTGGAGTCGTA 281
            ||||||||||||||||||| |||||| |||
Db      315 CCACTTATTACTGCCAGCATTGGAGTAGTA 344


RESULT    3
AI180569
LOCUS       AI180569         310 bp    mRNA             EST       08-OCT-1998
DEFINITION  uc70f09.r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone
            1431017 5' similar to gb:M18512 IG HEAVY CHAIN PRECURSOR V-I REGION
            (HUMAN); gb:M12376 Mouse immunoglobulin H-chain V-region pseudogene
            mRNA, complete (MOUSE);, mRNA sequence.
ACCESSION   AI180569
NID         g3731207
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 310)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:915085
            Seq primer: -28m13 rev2 ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .310
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
                     /db_xref="taxon:10090"
                     /clone="1431017"
                     /clone_lib="Soares mouse mammary gland NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
BASE COUNT       80 a     77 c     85 g     67 t      1 others
ORIGIN


  Query Match             24.4%;  Score 173.6;  DB 30;  Length 310;
  Best Local Similarity   79.8%;  Pred. No. 2.1e-44;
  Matches  217;  Conservative    0;  Mismatches   54;  Indels    1;  Gaps    1;

Qy    388 TGAGGTGGTGAAGCCTGGAGGTTCAATGAAGATATCCTGCAAGACTTCTGGTTACTCATT 447
          | || | ||||||||||| |||||| ||||| | ||||||||| ||||||| ||| | ||
Db     26 TAAGCTTGTGAAGCCTGG-GGTTCAGTGAAGCTGTCCTGCAAGGCTTCTGGCTACACCTT 84

Qy    448 CACTGGCCACACCATGAACTGGGTGAAGCAGAGCCATGGAAAGAACCTTGAGTGGATTGG 507
          |||  || ||   ||| | ||||| || ||||| | |||| ||  | |||||||||| ||
Db     85 CACCAGCTACTGGATGCAGTGGGTAAAACAGAGGCCTGGACAGGGCATTGAGTGGATCGG 144

Qy    508 ACTTATTAATCCTTACAATGGTGATACTAACTACAACCAGAAGTTCAAGGGCAAGGCCAC 567
          |   ||| ||||||   || |  ||||||||||||| || |||||||||||||||||  |
Db    145 AGAGATTGATCCTTCTGATAGCTATACTAACTACAATCAAAAGTTCAAGGGCAAGGCAGC 204

Qy    568 ATTTACTGTAGACAAGTCGTCCAGCACAGCCTACATGGAGCTCCTCAGTCTGACATCTGA 627
          ||| ||||||||||  || |||||||||||||||||| |||||  ||| |||||||||||
Db    205 ATTGACTGTAGACACATCCTCCAGCACAGCCTACATGCAGCTCAGCAGCCTGACATCTGA 264
```

```
Qy    628 GGACTCTGCAGTCTATTACTGTGCAAGGAGGG 659
          ||||||||| |||||| ||||||||||   ||
Db    265 GGACTCTGCGGTCTATGACTGTGCAAGCCAGG 296


RESULT   4
AI007196
LOCUS       AI007196      294 bp    mRNA            EST       12-JUN-1998
DEFINITION  ua73g04.r1 Soares 2NbMT Mus musculus cDNA clone 1363158 5' similar
            to gb:M18512 IG HEAVY CHAIN PRECURSOR V-I REGION (HUMAN); gb:M12376
            Mouse immunoglobulin H-chain V-region pseudogene mRNA, complete
            (MOUSE);, mRNA sequence.
ACCESSION   AI007196
NID         g3216753
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 294)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:896378
            Trace considered overall poor quality
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .294
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library went through two
                     rounds of normalization, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
                     /db_xref="taxon:10090"
                     /clone="1363158"
                     /clone_lib="Soares 2NbMT"
                     /sex="male"
                     /tissue_type="Thymus"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
BASE COUNT       85 a     68 c     77 g     64 t
ORIGIN


  Query Match             22.8%;  Score 162.4;  DB 28;  Length 294;
  Best Local Similarity   76.7%;  Pred. No. 7e-41;
  Matches  224;  Conservative    0;  Mismatches   66;  Indels    2;  Gaps    2;

Qy    365 GTGCAGCTGCAGGAGTCAGGACCTGAG-GTGGTGAAGCCTGGAGGTTCAATGAAGATATC 423
          || || |||||| || | ||  |||||  |||||| | ||||   |||| ||||| | |
Db       2 GTCCAACTGCAGCAGCCTGGGGCTGAGCATGGTGAGGACTGGGTCTTCAGTGAAGCTGTA 61

Qy     424 CTGCAAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAAGCAGAGCCA 483
           |||||||  ||| || ||| | |||||  || ||   ||| | |||||||||||||| ||
Db      62 CTGCAAGGATTCAGGCTACACCTTCACCAGCTACTGGATGCATTGGGTGAAGCAGAGGCA 121

Qy     484 TGGAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATACTAACTACAA 543
           |  | |   | |||| ||||||||    ||| | ||||   || |||  ||| |||||||
Db     122 TATACAAGGCATTGAATGGATTGGTAACATTGACCCTTCAGATAGTGGAACTCACTACAA 181

Qy     544 CCAGAAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCACAGCCTACA- 603
            || |||||||||| |||||||||||| ||||||||||| |   |||||||||||||||
Db     182 TCAAAAGTTCAAGGACAAGGCCACATTGACTGTAGACAAATTCACCAGCACAGCCTACAG 241

Qy     603 TGGAGCTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAG 654
           || |||||  ||| |||||||||||||||||||| |||||||||||||||||
Db     242 TGCAGCTCAGCAGACTGACATCTGAGGACTCTGCGGTCTATTACTGTGCAAG 293


RESULT   5
AI201426/c
LOCUS       AI201426      461 bp    mRNA            EST       30-OCT-1998
DEFINITION  qs73c06.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:1943722 3'
            similar to gb:M18512 IG HEAVY CHAIN PRECURSOR V-I REGION
            (HUMAN);contains Alu repetitive element;, mRNA sequence.
ACCESSION   AI201426
NID         g3754032
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            unknown library type
            Insert Length: 608   Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Homo sapiens"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr22 was prepared, and ss
                     circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     985608-986759, 1101192-1101959, and 1217928-1220615).
                     Subtraction by Bento Soares and M. Fatima Bonaldo.  "
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1943722"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
BASE COUNT       93 a    132 c    106 g    130 t
ORIGIN


  Query Match             22.8%;  Score 162;  DB 31;  Length 461;
  Best Local Similarity   72.4%;  Pred. No. 1.1e-40;
  Matches  210;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy     365 GTGCAGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTCAATGAAGATATCC 424
           |||||||||  | |||| ||  ||||||||  ||||||||| |  ||| ||||| | |||
Db     435 GTGCAGCTGGTGCAGTCTGGGGCTGAGGTGAAGAAGCCTGGGGCCTCAGTGAAGGTCTCC 376
```

```
Qy     425 TGCAAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAAGCAGAGCCAT 484
           |||||| ||||||| |||   ||||| | | ||   ||| ||||||||   |||  || |
Db     375 TGCAAGGCTTCTGGATACATCTTCACCGACTACTATATGCACTGGGTGCGACAGGCCCCT 316

Qy     485 GGAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATACTAACTACAAC 544
           ||| |  | |||| |||||| ||||  || || ||| ||| |||||  || |||||
Db     315 GGACAAGAGCTTGGGTGGATGGGACGGATCAACCCTAACAGTGGTGGCACAAACTATGCA 256

Qy     545 CAGAAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCACAGCCTACATG 604
           ||||||||  ||||||  | |||  | ||    |||| |||   |||||||||||||| |
Db     255 CAGAAGTTTCAGGGCAGAGTCACCATGACCAGGGACACGTCCATCAGCACAGCCTACACG 196

Qy     605 GAGCTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAG 654
           |||||   ||| |||| |||||||||| | ||    ||||||||||| ||
Db     195 GAGCTGAGCAGCCTGAGATCTGAGGACACGGCCACGTATTACTGTGCGAG 146


RESULT   6
AA569186
LOCUS       AA569186      387 bp    mRNA            EST       09-SEP-1997
DEFINITION  nm30d10.s1 NCI_CGAP_Lip2 Homo sapiens cDNA clone IMAGE:1061683
            similar to gb:L02325 IG HEAVY CHAIN PRECURSOR V-I REGION (HUMAN);,
            mRNA sequence.
ACCESSION   AA569186
NID         g2342240
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 387)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 1722   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 284.
FEATURES             Location/Qualifiers
     source          1. .387
                     /organism="Homo sapiens"
                     /note="Vector: pAMP10; mRNA made from liposarcoma, cDNA
                     made by oligo-dT priming. Non- directionally cloned.
                     Size-selected on agarose gel, average insert  size 600 bp.
                     Reference:  Krizman et al. (1996) Cancer Research
                     56:5380-5383."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1061683"
                     /clone_lib="NCI_CGAP_Lip2"
                     /tissue_type="liposarcoma"
                     /lab_host="DH10B"
     mRNA            <1. .>387
BASE COUNT       89 a     95 c    118 g     85 t
ORIGIN


  Query Match             20.5%;  Score 146;  DB 22;  Length 387;
  Best Local Similarity   69.0%;  Pred. No. 1.1e-35;
  Matches  200;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;
```

```
Qy      365 GTGCAGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTCAATGAAGATATCC 424
            || || |||  | |||| ||  ||||||||  ||||||||| |  ||| ||||| | |||
Db       98 GTCCAACTGGTGCAGTCTGGGGCTGAGGTGAAGAAGCCTGGGGCCTCAGTGAAGCTTTCC 157

Qy      425 TGCAAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAAGCAGAGCCAT 484
            || ||| | ||||| | | |  ||||   |  |   ||| ||||||||   |||  || |
Db      158 TGTAAGGCATCTGGATTCACCCTCACCAACTGCCATATGCACTGGGTGCGACAGGCCCCT 217

Qy      485 GGAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATACTAACTACAAC 544
            ||  |    ||||||||| | ||| | || ||| ||    ||||| |||  |   ||
Db      218 GGGCAAGGGCTTGAGTGGGTGGGAATGATCAATTCTAGTGATGGTTATATAAGTAACGCA 277

Qy      545 CAGAAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCACAGCCTACATG 604
            ||||||||| ||||||  | |||  | ||    |||| |||  | ||||||| ||| |||
Db      278 CAGAAGTTCCAGGGCAGAGTCACCATGACCAGGGACACGTCCACGAGCACAGTCTATATG 337

Qy      605 GAGCTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAG 654
            |||||   ||| |||| |||||||||| | || |||||||||||||| ||
Db      338 GAGCTTAGCAGCCTGAGATCTGAGGACACGGCCGTCTATTACTGTGCGAG 387


RESULT   7
AA710291
LOCUS       AA710291      345 bp    mRNA            EST       24-DEC-1997
DEFINITION  vt53a04.r1 Barstead mouse irradiated colon MPLRB7 Mus musculus cDNA
            clone 1166766 5' similar to gb:X02484_rna1 IG KAPPA CHAIN PRECURSOR
            V-IV REGION (HUMAN); gb:J00560 mouse ig kappa mrna from mopc21 &
            other myeloma mrna 3' (MOUSE);, mRNA sequence.
ACCESSION   AA710291
NID         g2720209
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 345)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:632678
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 140.
FEATURES             Location/Qualifiers
     source          1. .345
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; Tissue obtained
                     from 8 week old mouse.  Colon was harvested 72 hours after
                     irradiation with 1400 Gys.  1st strand cDNA was primed
                     with  a Not I - oligo(dT) primer
                     [5'TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTT
                     T 3'];  double-stranded cDNA was ligated to Eco RI
                     adaptors [AATTCGGATCCTTG],  digested with Not I and cloned
                     into the Not I and Eco RI sites of  the modified pT7T3
                     vector. Library constructed by Bob Barstead."
                     /db_xref="taxon:10090"
                     /clone="1166766"
```

```
                     /clone_lib="Barstead mouse irradiated colon MPLRB7"
                     /dev_stage="8 weeks"
                     /lab_host="DH10B"
BASE COUNT       95 a     81 c     88 g     81 t
ORIGIN


  Query Match             18.0%;  Score 128.4;  DB 25;  Length 345;
  Best Local Similarity   64.4%;  Pred. No. 3.8e-30;
  Matches  206;  Conservative    1;  Mismatches  112;  Indels    1;  Gaps    1;

Qy        4 GACCTGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTACAC 63
            ||| |   | |||||||||||| |  | || ||||  ||||  | || || | |||   |
Db       25 GACATTGTGATGACCCAGTCTCAAAAATTCATGTCCACATCAGCTGGAGACAGGGTCAGC 84

Qy       64 AATGACTTGCAGGGCCACCCCAAGT-GTAAGTTACATGCACTGGTATCAGCAGAAGCCAG 122
                 ||   ||     |   | || |  | | |  |   |||||||||  |||| ||||
Db       85 GTCACCTGCAAGCCAGTCAGAATGTGGGTACTAATGTAGCCTGGTATCAAGAGAAACCAG 144

Qy      123 GATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGTCCCTGCTC 182
            |    || || | | |   ||||||  |  ||||| ||| :     ||||||||||| ||
Db      145 GGCAATCTCCTAGAGCACTGATTTACTCGGCATCCTACCGGTACAGTGGAGTCCCTGATC 204

Qy      183 GCTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGAGGCTG 242
            ||||||  ||| |||| ||||||||   || | |||||||  ||||||  ||| || |||
Db      205 GCTTCACAGGCAGTGGATCTGGGACAGATTTCACTCTCACCATCAGCAATGTGCAGTCTG 264

Qy      243 AAGATGCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGGAGGGG 302
            ||||    ||    |||| ||| ||||| |  |   |    ||   ||||||||||||||
Db      265 AAGACTTGGCAGAGTATTTCTGTCAGCAATATAACAGCTTTCCGTACACGTTCGGAGGGG 324

Qy      303 GGTCCAAGCTGGAAATAAAA 322
            || |||||||||||||||||
Db      325 GGACCAAGCTGGAAATAAAA 344


RESULT   8
AA710970
LOCUS       AA710970      431 bp    mRNA            EST       24-DEC-1997
DEFINITION  vt93h09.r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone
            1178753 5' similar to gb:M18512 IG HEAVY CHAIN PRECURSOR V-I REGION
            (HUMAN); gb:X70423 M.musculus mRNA for monoclonal antibody heavy
            chain gamma (MOUSE);, mRNA sequence.
ACCESSION   AA710970
NID         g2720888
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 431)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:636601
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 198.
FEATURES             Location/Qualifiers
     source          1. .431
```

```
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
                     /db_xref="taxon:10090"
                     /clone="1178753"
                     /clone_lib="Soares mouse mammary gland NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
BASE COUNT      112 a    103 c    112 g    104 t
ORIGIN


  Query Match             16.7%;  Score 119;  DB 25;  Length 431;
  Best Local Similarity   62.0%;  Pred. No. 3.6e-27;
  Matches  188;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy    366 TGCAGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTCAATGAAGATATCCT 425
          | ||| ||    |||| ||||||||| ||  |||||||||||   || | ||||| ||||
Db    112 TCCAGTTGGTACAGTCTGGACCTGAGCTGAAGAAGCCTGGAGAGACAGTCAAGATCTCCT 171

Qy    426 GCAAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAAGCAGAGCCATG 485
          ||||| ||||||| ||  | |||||   | |    |||| ||||||||| |||   |  |
Db    172 GCAAGGCTTCTGGGTATACCTTCACAACCTATGGAATGAGCTGGGTGAAACAGGCTCCAG 231

Qy    486 GAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATACTAACTACAACC 545
          |||||    |  ||||||| ||    || ||  | |||  ||| |   | |  ||
Db    232 GAAAGGGTTTAAAGTGGATGGGCTGGATAAACACCTACTCTGGAGTGCCAACATATGCTG 291

Qy    546 AGAAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCACAGCCTACATGG 605
          |  | ||||||||   |    | ||  || | || |  ||  ||||||| |||||  ||
Db    292 ATGACTTCAAGGGACGGTTTGCCTTCTCTTTGGAAACCTCTGCCAGCACTGCCTATTTGC 351

Qy    606 AGCTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGGAGGGTTACGG 665
          || ||  ||  || | |  ||||||| | ||    |||| ||||||    |   || |
Db    352 AGATCAACAACCTCAAAAATGAGGACACGGCTACATATTTCTGTGCCCCAATACTTTCTA 411

Qy    666 ACT 668
          |||
Db    412 ACT 414


RESULT   9
T27593
LOCUS       T27593        393 bp    mRNA            EST       06-SEP-1995
DEFINITION  EST100653 Homo sapiens cDNA 5' end similar to immunoglobulin kappa
            light chain, V region (GB:L01279) (HT:3043).
ACCESSION   T27593
NID         g609691
KEYWORDS    EST.
SOURCE      human primer=M13 Reverse library=Human Pancreas.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 393)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.-W.,
            Clayton,R.A., Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            FitzGerald,L.M., FitzHugh,W.M., Fritchman,J.L., Geoghagen,N.S.M.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle Jr,P.S.,
            Kelley,J.M., Klimek,K.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M.,
            Merrick,J.M., Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T.,
            Pellegrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L.,
            Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R.,
            Weidman,J.F., Li,Y., Bednarik,D.P., Cao,L., Cepeda,M.A.,
            Coleman,T.A., Collins,E.-J., Dimke,D., Feng,P., Ferrie,A.,
            Fischer,C., Hastings,G.A., He,W.-W., Hu,J.-S., Greene,J.M.,
            Gruber,J., Hudson,P., Kim,A., Kozak,D.L., Kunsch,C., Ji,H., Li,H.,
            Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.-F., Wing,J., Xu,C.,
            Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A.,
            Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
  TITLE     Initial Assessment of Human Gene Diversity and Expression Patterns
            Based Upon 52 Million Basepairs of cDNA Sequence
  JOURNAL   Unpublished (1995)
COMMENT     Other_ESTs: THC24356
            Contact: Venter, JC
            The Institute for Genomic Research
            932 Clopper Rd, Gaithersburg, MD 20878
            Tel: 3018699056
            Fax: 3018699423
            Email: tdbinfo@tdb.tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please contact the TIGR Database
            (tdbinfo@tdb.tigr.org).
FEATURES             Location/Qualifiers
     source          1..393
                     /organism="Homo sapiens"
     mRNA            <1..>393
BASE COUNT       88 a    113 c     96 g     94 t      2 others
ORIGIN


  Query Match             16.7%;  Score 118.6;  DB 11;  Length 393;
  Best Local Similarity   65.7%;  Pred. No. 4.7e-27;
  Matches  216;  Conservative    1;  Mismatches  107;  Indels    5;  Gaps    3;

Qy      4 GACCTGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTACAC 63
          ||| | ||| |||||||||||||| |   |||||||||||||  ||| || |  || |||
Db     53 GACATCCAGATGACCCAGTCTCCATCCTCCCTGTCTGCATCTGTAGGAGACAGAGT-CAC 111

Qy     64 AATGACTTGCAGGGCCACCCCAAGTGTAAGTTACATGC---ACTGGTATCAGCAGAAGCC 120
            | |||||| |||| |  |  ||  | ||   | |     | |||||||||||||| ||
Db    112 CCTCACTTGCCGGGCAAGTCAGAGCATTAGCACCTTTTTAAATTGGTATCAGCAGAAACC 171

Qy    121 AGGATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGTCCCTGC 180
          |||  |  |||| || |    ||| | | |  |||||||  |:     ||| |||||  |
Db    172 AGGGACAGCCCCTAAGCTCCTGATCTCTGCTGCATCCAATTTGCGAAGTGGGGTCCCATC 231

Qy    181 TCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGAGGC 240
            | |||||||||  ||| ||||||||   || | |||||||  |||||||  || |  |
Db    232 AAGGTTCAGTGGCTCTGGATCTGGGACAGATTTCACTCTCACCATCAGCAGTCTGCAACC 291

Qy    241 TGAAGATGCTGCCACTTATTACTGCCAGCAGTG-GAGTCGTAGCCCACCCACGTTCGGAG 299
          |||||||| ||| ||||| ||||| || ||| |     | ||   | |  ||| | ||
Db    292 TGAAGATGTTGCAACTTACTACTGTCAACAGAGTTNCACTTATGNCTCGGACGGTTGGCC 351

Qy    300 GGGGGTCCAAGCTGGAAATAAAAGGTTCT 328
            ||| ||||| ||||  | ||| |   ||
Db    352 AAGGGACCAAGGTGGAGGTCAAACGGACT 380


RESULT  10
AA291381
LOCUS       AA291381      379 bp    mRNA            EST       08-AUG-1997
DEFINITION  zt44g02.r1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 725234
            5' similar to gb:M18512 IG HEAVY CHAIN PRECURSOR V-I REGION
            (HUMAN);, mRNA sequence.
ACCESSION   AA291381
NID         g1939359
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
```

```
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 379)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1576   Std Error: 0.00
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 361.
FEATURES             Location/Qualifiers
     source          1. .379
                     /organism="Homo sapiens"
                     /note="Organ: ovary; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGTTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library constructed by Bento Soares and
                     M.Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="725234"
                     /clone_lib="Soares ovary tumor NbHOT"
                     /sex="Female"
                     /tissue_type="ovarian tumor"
                     /lab_host="DH10B (ampicillin resistant)"
     mRNA            <1. .>379
BASE COUNT       85 a     82 c    121 g     91 t
ORIGIN


  Query Match             16.6%;  Score 118;  DB 21;  Length 379;
  Best Local Similarity   65.1%;  Pred. No. 7.3e-27;
  Matches  190;  Conservative    0;  Mismatches  100;  Indels    2;  Gaps    1;

Qy    363 GTGTGCAGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTCAATGAAGATAT 422
          | || ||| ||  | |||| ||||||||||||  |||||   | ||   | ||| | | |
Db     35 GGGTTCAGTTGGTGCAGTCTGGACCTGAGGTGAAGAAGC--TGGGGCCCAGTGAGGGTCT 92

Qy    423 CCTGCAAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAAGCAGAGCC 482
          |||| ||||| ||||||||| | ||             || |||||||||  ||||  ||
Db     93 CCTGTAAGACCTCTGGTTACACCTTTGGCACTTTTGGTATCAACTGGGTGCGGCAGGCCC 152

Qy    483 ATGGAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATACTAACTACA 542
           |||| |    ||||||||||| |||   || |    |||||||||| | || ||||
Db    153 CTGGACAAGGGCTTGAGTGGATGGGATGGATCAGCGTTTACAATGGTAAGACAAACTTTG 212

Qy    543 ACCAGAAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCACAGCCTACA 602
              ||  ||  ||| ||  | ||  || ||    ||||  ||  | |  ||||||||||
Db    213 CAGGGAGTTTTCAGGACAGAGTCAGCTTGACCACTGACACATCCACGAATACAGCCTACA 272

Qy    603 TGGAGCTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAG 654
          |||||||    |  |||| |||||| ||| |  |    ||||||||||| ||
Db    273 TGGAGCTGAGGAACCTGAGATCTGACGACACGGCCATATATTACTGTGCGAG 324


RESULT  11
AA318377
LOCUS       AA318377      335 bp    mRNA            EST       19-APR-1997
DEFINITION  EST20620 Spleen I Homo sapiens cDNA 5' end similar to similar to
            immunoglobulin kappa light chain, V region, anti-thyroglobulin
            (GB:X79786), mRNA sequence.
ACCESSION   AA318377
NID         g1970863
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 335)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .335
                     /organism="Homo sapiens"
                     /note="Organ: spleen; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI"
                     /db_xref="ATCC (inhost):119058"
                     /db_xref="taxon:9606"
                     /clone_lib="Spleen I"
                     /sex="male"
                     /dev_stage="adult, 23 yrs"
     mRNA            <1. .>335
BASE COUNT       87 a     88 c     71 g     84 t      5 others
ORIGIN


  Query Match             16.3%;  Score 116.2;  DB 17;  Length 335;
  Best Local Similarity   65.5%;  Pred. No. 2.6e-26;
  Matches  182;  Conservative    1;  Mismatches   92;  Indels    3;  Gaps    1;

Qy     61 CACAATGACTTGCAGGGCCACCCCAAG---TGTAAGTTACATGCACTGGTATCAGCAGAA 117
          ||| || |||||| |||| |  |  ||   |   || ||  | || ||||||||||| ||
Db      2 CACCATCACTTGCCGGGCAAGTCAGAGCATTAGCAGCTATTTACATTGGTATCAGCACAA 61

Qy    118 GCCAGGATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGTCCC 177
           |||||     |||| ||||    ||| |||   |||||||   |:     ||| |||||
Db     62 ACCAGGTAAAGCCCCTAAACTCCTGATCTATGNTACATCCACTTTGCAAAGTGGGGTCCC 121

Qy    178 TGCTCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGA 237
```

```
            |  | |||||||||  ||| ||||||||   || | |||||||  |||||||  || |
Db      122 ATCAAGGTTCAGTGGCACTGGATCTGGGACAGATTTCACTCTCACCATCAGCAGTCTGCA 181

Qy      238 GGCTGAAGATGCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGG 297
              ||||||||  ||| ||||| ||||| || ||| |     |||||||    ||||||||
Db      182 ACCTGAAGATTTTGCAACTTACTACTGTCAACAGAGTTCCAGTAGCCCTTGGACGTTCGG 241

Qy      298 AGGGGGGTCCAAGCTGGAAATAAAAGGTTCTACCTCTG 335
                ||| ||||| ||| ||| ||| |  ||    |||
Db      242 CCAAGGGNCCAAGGTGGCAATCAAACGNACTGTGGCTG 279


RESULT  12
AA301347
LOCUS       AA301347      413 bp    mRNA            EST       18-APR-1997
DEFINITION  EST14279 Testis tumor Homo sapiens cDNA 5' end similar to similar
            to immunoglobulin kappa light chain, V region (GB:L01279), mRNA
            sequence.
ACCESSION   AA301347
NID         g1953680
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .413
                     /organism="Homo sapiens"
                     /note="Organ: testis; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI"
                     /db_xref="ATCC (inhost):191658"
                     /db_xref="taxon:9606"
                     /clone_lib="Testis tumor"
                     /sex="male"
                     /dev_stage="adult"
     mRNA            <1. .>413
BASE COUNT      100 a    110 c    105 g     98 t
ORIGIN
```

```
  Query Match             16.0%;  Score 114.2;  DB 17;  Length 413;
  Best Local Similarity   65.0%;  Pred. No. 1.2e-25;
  Matches  214;  Conservative    1;  Mismatches  109;  Indels    5;  Gaps    3;

Qy        4 GACCTGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTACAC 63
            ||| | ||| |||||||||||||| |   |||||||||||||   || || |  || |||
Db       78 GACATTCAGATGACCCAGTCTCCATCCTCCCTGTCTGCATCTGTGGGAGACAGAGT-CAC 136

Qy       64 AATGACTTGCAGGGCCACCCCA---AGTGTAAGTTACATGCACTGGTATCAGCAGAAGCC 120
             || |||||| ||||    |     | |   |  ||  |  | |||||||| ||||| ||
Db      137 CATCACTTGCCGGGCTGGGCAGCACATTAACACCTATTTAAATTGGTATCAACAGAAACC 196

Qy      121 AGGATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGTCCCTGC 180
            |||     |||| || |    ||| |||    ||||||   |:     ||| |||||| |
Db      197 AGGGAAAGCCCCTAAGCTCCTGATGTATGGTGCATCCAGTTTGCAAGGTGGGGTCCCTTC 256

Qy      181 TCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGAGGC 240
              | |||||||||||||| ||||||||   || | |||||||  ||| |||  || |  |
Db      257 AAGGTTCAGTGGCGGTGGATCTGGGACAGATTTCACTCTCACCATCACCAGTCTGCAACC 316

Qy      241 TGAAGATGCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCAC-CCACGTTCGGAG 299
            |||||||  ||| ||||| ||||  |||||| |     ||   |  |   | ||||||
Db      317 TGAAGATTTTGCAACTTACTACTTTCAGCAGAGTTACAGTCCTCATCAGAAGGTTCGGCC 376

Qy      300 GGGGGTCCAAGCTGGAAATAAAAGGTTCT 328
              ||| ||||| ||||||| ||| |  ||
Db      377 AAGGGACCAAGGTGGAAATCAAAAGGACT 405


RESULT  13
AA987559
LOCUS       AA987559      336 bp    mRNA            EST       23-JUL-1998
DEFINITION  or83g01.s1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1602480 3'
            similar to gb:M87789 IG GAMMA-1 CHAIN C REGION (HUMAN);, mRNA
            sequence.
ACCESSION   AA987559
NID         g3172923
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 336)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 801   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 322.
FEATURES             Location/Qualifiers
     source          1. .336
                     /organism="Homo sapiens"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
                     neuroendocrine lung carcinoid, and was then primed with a
                     Not I - oligo(dT) primer. Double-stranded cDNA was ligated
                     to Eco RI adaptors (Pharmacia), digested with Not I and
```

```
                cloned into the Not I and Eco RI sites of the modified
                pT7T3 vector.  Library is normalized. Library was
                constructed by Bento Soares and M. Fatima Bonaldo."
                /db_xref="taxon:9606"
                /clone="IMAGE:1602480"
                /clone_lib="NCI_CGAP_Lu5"
                /tissue_type="carcinoid"
                /lab_host="DH10B"
BASE COUNT       82 a     86 c    100 g     68 t
ORIGIN

  Query Match              15.8%;  Score 112.4;  DB 29;  Length 336;
  Best Local Similarity    61.7%;  Pred. No. 4.1e-25;
  Matches  179;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy    365 GTGCAGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTCAATGAAGATATCC 424
          || ||| |  |  |||||||  ||||||||  |||||| ||   |||  ||||| |  ||
Db     32 GTCCAGATAGAACAGTCAGGGGCTGAGGTGAAGAAGCCAGGGTCTTCGGTGAAGGTCACC 91

Qy    425 TGCAAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAAGCAGAGCCAT 484
          |||  | | |||||   | | |||  | |    ||  | | |||||||   |||  || |
Db     92 TGCTTGGCCTCTGGAGGCACCTTCGGTAGTTCGACTGTCAGCTGGGTGCGACAGGCCCCT 151

Qy    485 GGAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATACTAACTACAAC 544
          ||| |    | ||||||||| ||    || |  ||  |   |  |    | ||||||
Db    152 GGACAACGGCCTGAGTGGATGGGGAAGATCATCCCCATCGCTCATTCAGCGAACTACGCA 211

Qy    545 CAGAAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCACAGCCTACATG 604
          | ||||||  ||| ||    |||| | || |  ||||| ||  |     ||| |||||||
Db    212 CCGAAGTTTCAGGACAGAATCACAATAACCGCGGACAAATCTACGGCGTCAGTCTACATG 271

Qy    605 GAGCTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAG 654
          ||  |   ||| |||| || ||||||| | ||  | ||||| ||||| ||
Db    272 GAATTGACCAGCCTGATATTTGAGGACACGGCCATATATTATTGTGCGAG 321


RESULT  14
AA345486
LOCUS       AA345486       403 bp    mRNA             EST       21-APR-1997
DEFINITION  EST51505 Gall bladder II Homo sapiens cDNA 5' end similar to
            similar to immunoglobulin kappa light chain, V region (GB:X06763),
            mRNA sequence.
ACCESSION   AA345486
NID         g1997722
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 403)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
```

```
COMMENT
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .403
                     /organism="Homo sapiens"
                     /note="Organ: gall bladder; Vector: pBluescript SK-;
                     Site_1: EcoRI; Site_2: XhoI"
                     /db_xref="ATCC (inhost):147008"
                     /db_xref="taxon:9606"
                     /clone_lib="Gall bladder II"
                     /sex="female"
                     /dev_stage="adult, 25 yrs"
     mRNA            <1. .>403
BASE COUNT       86 a    113 c    107 g     93 t      4 others
ORIGIN

  Query Match              15.6%;  Score 110.8;  DB 18;  Length 403;
  Best Local Similarity    63.1%;  Pred. No. 1.4e-24;
  Matches  188;  Conservative    1;  Mismatches  103;  Indels    6;  Gaps    1;

Qy     44 CTCCAGGGGAGAAGGTACACAATGACTTGCAGGGCCACCCCAAGTGT------AAGTTAC 97
          ||||  |||  |||   | |  |  | ||||||||||  |   ||||       ||| ||
Db     39 CTCCCAGGGGAAAGAGCCGCCCTCTCCTGCAGGGCCAGTCAGCGTGTTAGCAGCAGTCAC 98

Qy     98 ATGCACTGGTATCAGCAGAAGCCAGGATCCTCCCCCAAACCTTGGATTTATACCACATCC 157
           |   ||||||||||||||| || ||     | ||||  |     || |||    | |||
Db     99 TTAGCCTGGTATCAGCAGAAACCTGGCCAGGCTCCCAGGCTCCTCATCTATGGTGCGTCC 158

Qy    158 AACCTKGCTTCTGGAGTCCCTGCTCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCT 217
          |    :||  ||||  |||| |   | ||||||||| |||||||||||||    | | ||
Db    159 AGGAGGGCCACTGGCATCCCAGACAGGTTCAGTGGCAGTGGGTCTGGGACAGACTTCACT 218

Qy    218 CTCACAGTCAGCAGAGTGGAGGCTGAAGATGCTGCCACTTATTACTGCCAGCAGTGGAGT 277
          |||||  |||||||| ||||| ||||||||  |||    |||||||| |||||||   ||
Db    219 CTCACCATCAGCAGACTGGAGCCTGAAGATTTTGCAGTGTATTACTGTCAGCAGTATGGT 278

Qy    278 CGTAGCCCACCCACGTTCGGAGGGGGGTCCAAGCTGGAAATAAAAGGTTCTACCTCTG 335
                ||    ||||||||    ||| ||||| ||||||| ||| |  ||    |||
Db    279 GACTCACCTAGGACGTTCGGCCAAGGGACCAAGGTGGAAATCAAACGAACTGTGGCTG 336


RESULT  15
AA423447
LOCUS       AA423447       398 bp    mRNA             EST       16-OCT-1997
DEFINITION  ve80a03.r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone
            832492 5' similar to gb:M63438 IG KAPPA CHAIN PRECURSOR V-III
            REGION (HUMAN); gb:M35669 Mouse Ig aberrantly rearranged
            kappa-chain mRNA V-J2-C-region, (MOUSE);, mRNA sequence.
ACCESSION   AA423447
NID         g2102360
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 398)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:492708
            Seq primer: -28m13 rev2 ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .398
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
                     /db_xref="taxon:10090"
                     /clone="832492"
                     /clone_lib="Soares mouse mammary gland NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
BASE COUNT     101 a    100 c    102 g     95 t
ORIGIN


  Query Match             15.6%;  Score 110.8;  DB 24;  Length 398;
  Best Local Similarity   63.2%;  Pred. No. 1.4e-24;
  Matches  211;  Conservative    0;  Mismatches  107;  Indels   16;  Gaps    2;

Qy        4 GACCTGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTACAC 63
            ||| |   ||||||||| ||||||||   | || |||  |||| |||| || |||  |||
Db       66 GACATTGTGCTGACCCAATCTCCAGCTTCCTTGGCTGTGTCTCTAGGGCAG-AGGGCCAC 124

Qy       64 AATGACTTGCAGGGCCACCCCAAGTGTA---------------AGTTACATGCACTGGTA 108
             ||  | ||||| |||| |  ||||||                ||||  ||| ||||||
Db      125 CATCTCCTGCAGAGCCAGCGAAAGTGTTGATAATTCTGGCATTAGTTTTATGAACTGGTT 184

Qy      109 TCAGCAGAAGCCAGGATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTC 168
             || ||||| ||||||    | |||||||     || ||| |  ||||||||      ||
Db      185 CCAACAGAAACCAGGACAGGCACCCAAACTCCTCATCTATGCTGCATCCAACAAAAGATC 244

Qy      169 TGGAGTCCCTGCTCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAG 228
             || ||||||||  | || |||||| |||||||||||||    | |   ||||   ||
Db      245 CGGGGTCCCTGCCAGGTTTAGTGGCAGTGGGTCTGGGACAGACTTCAGCCTCAACATCCA 304

Qy      229 CAGAGTGGAGGCTGAAGATGCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACC 288
                 ||||||  || ||| || | |  |||| ||| |||||  | |        || |
Db      305 TCCTATGGAGGAGGATGATACTACAATGTATTTCTGTCAGCAAAGTAAGGAGGTTCCTCG 364

Qy      289 CACGTTCGGAGGGGGGTCCAAGCTGGAAATAAAA 322
             |||||||| || ||  ||||| ||||||| |||
Db      365 GACGTTCGGTGGAGGCACCAAGTTGGAAATCAAA 398


Search completed: May 13, 1999, 09:40:43
Job time: 361 sec
```

```
                      GenCore version 4.5
          Copyright (c) 1993 - 1998  Compugen Ltd.


OM nucleic - nucleic search, using sw model

Run on:          May 13, 1999, 09:35:42 ; Search time 59.56 Seconds
                                          (without alignments)
                                          2248.983 Million cell updates/sec

Title:           US-08-704-178-1
Perfect score:   712
Sequence:        1 ATGGACCTGCAGCTGACCCA..........GGACCACGGTCACCGTCTCC 712

Scoring table:   IDENTITY_NUC

Searched:        240622 seqs, 94065609 residues

Database :       N_Geneseq_34:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                              SUMMARIES
                    %
Result            Query
  No.    Score    Match Length DB  ID                    Description
 -----------------------------------------------------------------------
     1   699.6    98.3    711  1  T65006                 Single-chain anti-
     2   696.4    97.8    711  1  Q55180                 Sequence encoding
     3   694.8    97.6    711  1  T17728                 Anti-erbB2 scFv cD
     4     458    64.3    720  1  T91615                 cDNA encoding an a
     5     434    61.0    732  1  Q73678                 Fv(GP-2) immunosup
     6     422    59.3    720  1  Q55181                 Sequence encoding
     7     422    59.3    720  1  T65007                 Single-chain anti-
     8   356.2    50.0    913  1  Q81076                 Bispecific CD3-L6F
     9   350.4    49.2    906  1  V00611                 Anti-human SC sing
    10   348.4    48.9    729  1  Q73679                 Fv(GP-4) immunosup
    11   331.6    46.6    848  1  Q41069                 26-10 sFv coding s
    12   320.6    45.0    738  1  Q12405                 OVB3 light and hea
    13     319    44.8    797  1  T29058                 CC49/212 SCA PLAP
    14   317.8    44.6   1460  1  Q43291                 Sequence encoding
    15   317.8    44.6    803  1  T29059                 PLAP CC49/212 SCA
    16   316.6    44.5    731  1  Q43287                 Sequence encoding
    17   315.8    44.4    735  1  T91614                 cDNA encoding an a
    18   313.2    44.0    741  1  Q51538                 Coding sequence of
    19   313.2    44.0    741  1  T13737                 Single chain bindi
    20   313.2    44.0    741  1  T36461                 TRY59 single chain
    21   311.6    43.8    741  1  Q05712                 TRY59. Single poly
    22   308.8    43.4    753  1  Q05713                 TRY104b. Single po
    23   308.8    43.4    753  1  Q51539                 Coding sequence of
    24   308.8    43.4    753  1  T13738                 Single chain bindi
    25   308.8    43.4    753  1  T36462                 TRY104b single cha
    26   298.6    41.9    724  1  Q05714                 18-2-3-/TRY202'. S
    27   298.6    41.9    720  1  Q51540                 Coding sequence of
    28   298.6    41.9    720  1  T13739                 Single chain bindi
    29   298.6    41.9    720  1  T36463                 18-2-3/TRY202' sin
    30   296.6    41.7    819  1  Q20381                 Sequence encoding
    31     293    41.2    725  1  Q66840                 4-4-20 VL/ 217 / C
    32   291.4    40.9    725  1  Q43292                 Sequence encoding
    33   287.2    40.3   6799  1  V63619                 Nucleic acid seque
    34   285.8    40.1    819  1  T35691                 Nematode salivary
    35   285.8    40.1    819  1  T86646                 Mus musculus antib
    36   285.2    40.1    729  1  Q51541                 Coding sequence of
    37   285.2    40.1    729  1  T13740                 Single chain bindi
    38   285.2    40.1    729. 1  T36464                 18-2-3/TRY59 singl
    39   285.2    40.1   6727  1  V63616                 Nucleic acid seque
    40   284.2    39.9   1320  1  V63617                 Nucleic acid seque
    41   284.2    39.9   1299  1  V63620                 Nucleic acid seque
    42   282.8    39.7    836  1  Q45604                 Sequence of single
    43   282.2    39.6    462  1  T31541                 3H1 heavy chain va
    44   282.2    39.6    462  1  T34542                 Monoclonal anti-id
    45   282.2    39.6    462  1  T99435                 Anti-idiotype anti


                              ALIGNMENTS


RESULT   1
T65006
ID   T65006 standard; cDNA; 711 BP.
AC   T65006;
DT   05-JUN-1997  (first entry)
DE   Single-chain anti-erbB2 antibody e23(Fv) cDNA.
KW   Single chain antibody; variable region; light chain; heavy chain;
KW   breast cancer; ovarian cancer; non-small cell lung carcinoma;
KW   immunodiagnosis; treatment; cytotoxic agent; erbB-2; ds.
OS   Mus musculus.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   mat_peptide     1. .711
FT                   /*tag= a
FT                   /product= e23(Fv)
PN   US5587458-A.
PD   24-DEC-1996.
PF   07-OCT-1991; 772270.
PR   07-OCT-1991; US-772270.
PR   30-JUN-1992; US-906555.
PR   14-MAY-1993; US-061092.
PA   (ARON-) ARONEX PHARM INC.
PI   Bird RE,  Kasprzyk PG,  King CR;
DR   WPI; 97-064831/06.
DR   P-PSDB; W15185.
PT   Single chain antibodies specific for erbB-2 protein, gp185  - with
PT   labels or cytotoxin, useful for detection and treatment of tumour
PT   cells expressing this protein
PS   Example 8; Columns 25-28; 28pp; English.
CC   The present cDNA sequence codes for a claimed single-chain antibody,
CC   designated e23(Fv), which binds to erbB-2. Monoclonal antibody e23
CC   was generated by immunising mice with N/erbB-2 cells overexpressing
CC   the gp185 protein, removing spleen cells and producing hybridomas
CC   by standard techniques. Messenger RNA coding for the anti-erbB-2
CC   monoclonal antibody was isolated and converted to cDNA. Regions
CC   coding for the heavy- and light- chain variable regions were then
CC   amplified by PCR and joined via a sequence encoding a peptide
CC   linker. The resulting single-chain antibody is useful for in vitro
CC   diagnosis of tumour cells which overexpress the erbB-2 gp185
CC   marker, e.g. breast, ovarian and non-small cell lung carcinomas,
CC   and, when coupled to a cytotoxic agent, to treat such tumours.
SQ   Sequence   711 BP;    175 A;    182 C;    189 G;    164 T;


  Query Match             98.3%;  Score 699.6;  DB 1;  Length 711;
  Best Local Similarity   99.9%;  Pred. No. 4.7e-177;
  Matches  711;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy      1 ATGGACCTGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db      1 ATGGACCTGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGT- 60

Qy     61 CACAATGACTTGCAGGGCCACCCCAAGTGTAAGTTACATGCACTGGTATCAGCAGAAGCC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60 CACAATGACTTGCAGGGCCACCCCAAGTGTAAGTTACATGCACTGGTATCAGCAGAAGCC 119

Qy    121 AGGATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGTCCCTGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    120 AGGATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGTCCCTGC 179

Qy    181 TCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGAGGC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    180 TCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGAGGC 239

Qy    241 TGAAGATGCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGGAGG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     240 TGAAGATGCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGGAGG 299

Qy     301 GGGGTCCAAGCTGGAAATAAAAGGTTCTACCTCTGGTTCTGGTAAATCTTCTGAAGGTAA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     300 GGGGTCCAAGCTGGAAATAAAAGGTTCTACCTCTGGTTCTGGTAAATCTTCTGAAGGTAA 359

Qy     361 AGGTGTGCAGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTCAATGAAGAT 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     360 AGGTGTGCAGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTCAATGAAGAT 419

Qy     421 ATCCTGCAAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAAGCAGAG 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     420 ATCCTGCAAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAAGCAGAG 479

Qy     481 CCATGGAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATACTAACTA 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     480 CCATGGAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATACTAACTA 539

Qy     541 CAACCAGAAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCACAGCCTA 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     540 CAACCAGAAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCACAGCCTA 599

Qy     601 CATGGAGCTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGGAGGGT 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     600 CATGGAGCTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGGAGGGT 659

Qy     661 TACGGACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCC 712
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     660 TACGGACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCC 711


RESULT   2
Q55180
ID   Q55180 standard; cDNA; 711 BP.
AC   Q55180;
DT   21-JUL-1994  (first entry)
DE   Sequence encoding the single chain anti-erbB2 antibody, Ab no.23.
KW   Single chain anti-erbB1 antibody; cancer therapy; prevention;
KW   monoclonal antibody; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   cds             1. .171
FT                   /*tag=  a
PN   WO9400136-A.
PD   06-JAN-1994.
PF   21-OCT-1992; U08545.
PR   30-JUN-1992; US-906555.
PA   (MOLE-) MOLECULAR ONCOLOGY INC.
PI   Kasprzyk PG,  King CR;
DR   WPI; 94-025878/03.
DR   P-PSDB; R45442.
PT   Treatment of malignancies over-expressing ERB-[2 - using at least
PT   2 monoclonal antibodies which recognise different epitopes on
PT   gp185
PS   Example; Fig 7; 37pp; English.
CC   The source of human erbB-2 protein for the prodn. of antibodies no.
CC   23 and 21 (Q55180/R45442; Q55181/R45443) is a NIH/3T3 cell
CC   engineered to express the human erbB-2 protein on its surface (N/
CC   erbB-2). Abs no. 21 and 23 are directed against the extracellular
CC   domain of gp185 erbB-2. Nude mice manipulated to produce rapidly
CC   growing tumours were used in a trial of the efficacy of the Abs. In
CC   animals given a combination of the 2 Abs, tumours completely
CC   regressed after 11 days.
SQ   Sequence   711 BP;    175 A;    183 C;    187 G;    165 T;


  Query Match             97.8%;  Score 696.4;  DB 1;  Length 711;
  Best Local Similarity   99.6%;  Pred. No. 3.3e-176;
  Matches  709;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy       1 ATGGACCTGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTA 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGGACCTGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGT- 60
           •

Qy      61 CACAATGACTTGCAGGGCCACCCCAAGTGTAAGTTACATGCACTGGTATCAGCAGAAGCC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60 CACAATGACTTGCAGGGCCACCCCAAGTGTAAGTTACATGCACTGGTATCAGCAGAAGCC 119

Qy     121 AGGATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGTCCCTGC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 AGGATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGTCCCTGC 179

Qy     181 TCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGAGGC 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180 TCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGAGGC 239

Qy     241 TGAAGATGCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGGAGG 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     240 TGAAGATGCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGGAGG 299

Qy     301 GGGGTCCAAGCTGGAAATAAAAGGTTCTACCTCTGGTTCTGGTAAATCTTCTGAAGGTAA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     300 GGGGTCCAAGCTGGAAATAAAAGGTTCTACCTCTGGTTCTGGTAAATCTTCTGAAGGTAA 359

Qy     361 AGGTGTGCAGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTCAATGAAGAT 420
           |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db     360 AGGTGTGCAGCTGCAGGATTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTCAATGAAGAT 419

Qy     421 ATCCTGCAAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAAGCAGAG 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     420 ATCCTGCAAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAAGCAGAG 479

Qy     481 CCATGGAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATACTAACTA 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     480 CCATGGAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATACTAACTA 539

Qy     541 CAACCAGAAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCACAGCCTA 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     540 CAACCAGAAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCACAGCCTA 599

Qy     601 CATGGAGCTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGGAGGGT 660
           ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     600 CATGGAGCTCCTCAGTCTGACATCTGAGGACTCTGCACTCTATTACTGTGCAAGGAGGGT 659

Qy     661 TACGGACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCC 712
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     660 TACGGACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCC 711


RESULT   3
T17728
ID   T17728 standard; cDNA; 711 BP.
AC   T17728;
DT   21-MAY-1996  (first entry)
DE   Anti-erbB2 scFv cDNA.
KW   Oncoprotein; erbB2; cell proliferation; tumour; cancer;
KW   intracellular antibody homologue; single chain antibody; scFv;
KW   gene therapy; ds.
OS   Synthetic.
PN   WO9607321-A1.
PD   14-MAR-1996.
PF   23-AUG-1995; U10740.
PR   06-SEP-1994; US-301339.
PR   06-JUN-1995; US-468252.
PA   (UABR-) UAB RES FOUND.
PI   Curiel DT,  Deshane J;
DR   WPI; 96-171307/17.
DR   P-PSDB; R94020.
PT   Inhibition of proliferation or survival of, esp. malignant erbB2,
PT   cells - by introducing nucleic acid mol. encoding antibody homologue
PT   which is expressed and binds, pref. erbB2, protein intracellularly
PS   Claim 42; Page 29-30; 48pp; English.
CC   A nucleic acid comprises a first sequence encoding a signal peptide
CC   (R94019) linked to a second sequence (T17728) encoding a single
CC   chain Fv fragment (R94020) that binds a human erbB2 oncoprotein.
CC   The anti-erbB2 sFv portion is obtained by PCR using e23scFv
CC   plasmid as template.  The signal peptide directs the scFv to the
CC   endoplasmic reticulum.  The nucleic acid is incorporated into a
CC   plasmid or viral vector to facilitate expression of the scFv antibody
CC   homologue within e.g. an epithelial carcinoma cell.  Intracellular
CC   expression of the homologue inhibits surface expression of erbB2 and
CC   thereby inhibits cell proliferation and cell survival and decreases
CC   tumorigenicity.
SQ   Sequence   711 BP;    175 A;    183 C;    189 G;    164 T;


  Query Match             97.6%;  Score 694.8;  DB 1;  Length 711;
  Best Local Similarity   99.3%;  Pred. No. 8.9e-176;
  Matches  707;  Conservative    1;  Mismatches    3;  Indels    1;  Gaps    1;

Qy      1 ATGGACCTGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTA 60
          ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db      1 ATGGACCTGCAGCTGACCCAGTCTCCAGCAATCCTCTCTGCATCTCCAGGGGAGAAGGT- 60

Qy     61 CACAATGACTTGCAGGGCCACCCCAAGTGTAAGTTACATGCACTGGTATCAGCAGAAGCC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60 CACAATGACTTGCAGGGCCACCCCAAGTGTAAGTTACATGCACTGGTATCAGCAGAAGCC 119

Qy    121 AGGATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGTCCCTGC 180
          ||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||
Db    120 AGGATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTGGCTTCTGGAGTCCCTGC 179

Qy    181 TCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGAGGC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    180 TCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGAGGC 239

Qy    241 TGAAGATGCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGGAGG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    240 TGAAGATGCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGGAGG 299

Qy    301 GGGGTCCAAGCTGGAAATAAAAGGTTCTACCTCTGGTTCTGGTAAATCTTCTGAAGGTAA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    300 GGGGTCCAAGCTGGAAATAAAAGGTTCTACCTCTGGTTCTGGTAAATCTTCTGAAGGTAA 359

Qy    361 AGGTGTGCAGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTCAATGAAGAT 420
          |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db    360 AGGTGTGCAGCTGCAGGAGTCAGGACCTGAGGTGGTCAAGCCTGGAGGTTCAATGAAGAT 419

Qy    421 ATCCTGCAAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAAGCAGAG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    420 ATCCTGCAAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAAGCAGAG 479

Qy    481 CCATGGAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATACTAACTA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    480 CCATGGAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATACTAACTA 539

Qy    541 CAACCAGAAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCACAGCCTA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    540 CAACCAGAAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCACAGCCTA 599

Qy    601 CATGGAGCTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGGAGGGT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    600 CATGGAGCTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGGAGGGT 659

Qy    661 TACGGACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCC 712
          ||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    660 TACGGACTGGTACTTCGATGTCTGGGGGGCAGGGACCACGGTCACCGTCTCC 711


RESULT   4
T91615
ID   T91615 standard; cDNA to mRNA; 720 BP.
AC   T91615;
DT   22-DEC-1997  (first entry)
DE   cDNA encoding an anti-gp130 antibody derived scFv.
KW   Single chain Fv; scFv; antibody; E. coli; anti-T3 antibody;
KW   anti-gp130 antibody; inclusion body; chaperonin; ds.
OS   Synthetic.
PN   J09220092-A.
```

```
PD   26-AUG-1997.
PF   15-FEB-1996; 027622.
PR   15-FEB-1996; JP-027622.
PA   (TOYJ ) TOSOH CORP.
DR   WPI; 97-474306/44.
DR   P-PSDB; W25784.
PT   Producing single chain Fv antibody in Escherichia coli - by
PT   expression in an inclusion body, followed by protein folding or by
PT   co-expression with a chaperonin as a soluble fraction
PS   Claim 4; Page 6-7; 9pp; Japanese.
CC   The sequences given in T91614-15 encode single chain Fv (scFv)
CC   antibodies which are produced in E. coli.  The scFV's are derived
CC   from an anti-T3 antibody or an anti-gp130 antibody and are produced
CC   either by: (1) expression as an inclusion body, followed by folding
CC   (i.e. by denaturation and solubilisation) or (2) expression as a
CC   soluble fraction by co-expression with chaperonin.  Using the methods,
CC   scFv can be produced in E. coli, either in a soluble fraction or in
CC   inclusion bodies.  Like the parent double chain antibodies, the scFv
CC   polypeptides are useful in immunological diagnosis or for separation
CC   and purification, but they lessen the side effects caused by antibody
CC   constant regions.  The availability of scFv polypeptides contributes
CC   to fundamental research and development of diagnostic and therapeutic
CC   drugs.
SQ   Sequence   720 BP;    169 A;    193 C;    190 G;    168 T;


  Query Match             64.3%;  Score 458;  DB 1;  Length 720;
  Best Local Similarity   80.2%;  Pred. No. 6.4e-113;
  Matches  576;  Conservative    1;  Mismatches  131;  Indels   10;  Gaps    3;

Qy      4 GACCTGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTACAC 63
          || ||   ||| ||||||||||||||||| | |||||||||||||||||||||||| |||
Db      1 GAGCTCGTGCTCACCCAGTCTCCAGCATTGATATCTGCATCTCCAGGGGAGAAGGT-CAC 59

Qy     64 AATGACTTGCAGGGCCACCCCAAGTGTAAGTTACATGCACTGGTATCAGCAGAAGCCAGG 123
           ||||| ||||  | || | ||||||| | || |||| | ||||| ||||||||||||||
Db     60 CATGACCTGCAATGTCAGCTCAAGTGTTACTTCCATGTATTGGTACCAGCAGAAGCCAGG 119

Qy    124 ATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGTCCCTGCTCG 183
          ||||||||||||||| |||||||||  ||||||||||||:||||||||||||||| ||||
Db    120 ATCCTCCCCCAAACCCTGGATTTATCTCACATCCAACCTGGCTTCTGGAGTCCCTCCTCG 179

Qy    184 CTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGAGGCTGA 243
          || ||||||| |||||||||||||| |||||||||||||| |||||||  ||||||||||
Db    180 CTCCAGTGGCAGTGGGTCTGGGACCACTTACTCTCTCACAATCAGCAGCTTGGAGGCTGA 239

Qy    244 AGATGCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGGAGGGGG 303
          ||||||||||||||||||||||||||||||||||  || ||| | ||||||||| |  ||
Db    240 AGATGCTGCCACTTATTACTGCCAGCAGTGGAGTACTAACCCGCTCACGTTCGGTGCTGG 299

Qy    304 GTCCAAGCTGGAAATAAAAGGTTCTACCTCTGGTTCTGGTAAATCTTCTGAAGGTAAA-- 362
          | ||||||| ||  |    ||   |  ||| ||   ||||   ||   ||  ||   |
Db    300 GACCAAGCTCGAGCTGGGTGGCGGTGGCTCGGGCGGTGGTGGGTCGGGTGGCGGCGGATC 359

Qy    362 -GGTGTGCAGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTCAATGAAGAT 420
           |  || ||  ||||| |||| ||||||||  ||||||||||||| | |||| |||||||
Db    360 GGACGTCCAATTGCAGCAGTCTGGACCTGAACTGGTGAAGCCTGGGGCTTCAGTGAAGAT 419

Qy    421 ATCCTGCAAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAAGCAGAG 480
          | |||||||| |||| || ||| ||||||||| | ||| |||| ||||||||||||||||
Db    420 ACCCTGCAAGGCTTCAGGATACACATTCACTGACTACAACATGGACTGGGTGAAGCAGAG 479

Qy    481 CCATGGAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATACTAACTA 540
          ||||||||||| ||||||||||||||||  ||||||| || | | |||||  || || |||
Db    480 CCATGGAAAGAGCCTTGAGTGGATTGGAGATATTAATTCTCATAGTGGTGGTATTATCTA 539

Qy    541 CAACCAGAAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCACAGCCTA 600
          |||||| |||||||||| |||||||| ||| |||||||| ||||| ||||||||||||||
Db    540 CAACCAAAAGTTCAAGGACAAGGCCATATTGACTGTAGATAAGTCCTCCAGCACAGCCTA 599

Qy    601 CATGGAGCTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGGA---- 657
          ||||||||||| ||| |||||||||||||||| |||||||||||||| ||||||||| |
Db    600 CATGGAGCTCCGCAGCCTGACATCTGAGGACACTGCAGTCTATTATTGTGCAAGAACCTA 659
```

```
Qy        657 --GGGTTACGGACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCC 712
                  ||    | |   | ||| ||   |||||||  ||| |||||  |||| ||||||
Db        660 CTATGTGTACGGCCACTTCTTTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCC 717


RESULT   5
Q73678
ID   Q73678 standard; DNA; 732 BP.
AC   Q73678;
DT   09-MAY-1995  (first entry)
DE   Fv(GP-2) immunosuppressive.
KW   MAb; monoclonal antibody; hybridoma; interleukin-2; IL-2;
KW   Fv; antibody variable region; GP-2; Fv(GP-2);
KW   immunosuppressive; ss.
OS   Mus sp.
PN   EP-621338-A.
PD   26-OCT-1994.
PF   21-APR-1994; 106257.
PR   21-APR-1993; JP-094491.
PR   07-MAR-1994; JP-036065.
PA   (AJIN ) AJINOMOTO KK.
PI   Hamura J,  Kanayama Y,  Nakazawa H,  Shimamura T,  Sugamura K;
PI   Takeshita T;
DR   WPI; 94-325948/41.
DR   P-PSDB; R60780.
PT   Immunosuppressant polypeptide - has ability to block
PT   interleukin-2 response
PS   Claim 17; Page 29; 37pp; English.
CC   MAb capable of binding to the gamma chain of the IL-2 receptor, and
CC   thus of blocking the IL-2 response, is produced by mouse hybridoma
CC   line GP-2 (FERM BP-4641).  DNA encoding the variable region of
CC   this MAb was expressed in E. coli, yielding Fv(GP-2) with
CC   immumosuppressive activity.
SQ   Sequence   732 BP;    180 A;    204 C;    182 G;    166 T;


  Query Match             61.0%;  Score 434;  DB 1;  Length 732;
  Best Local Similarity   78.6%;  Pred. No. 1.5e-106;
  Matches  574;  Conservative    1;  Mismatches  136;  Indels   19;  Gaps    4;

Qy          1 ATGGACCTGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTA 60
              |||||| | | |||||||||||||||| ||||| |||||||||||| |||||| | ||
Db          1 ATGGACATCCTGCTGACCCAGTCTCCATCAATCATGTCTGCATCTCTAGGGGA-ACGGGT 59

Qy         61 CACAATGACTTGCAGGGCCACCCCAAGTGTA------AGTTACATGCACTGGTATCAGCA 114
              ||| ||||| ||||  |||| | ||||||||      |||||| |||||||||| |||||
Db         60 CACCATGACCTGCACTGCCAGCTCAAGTGTAAGTTCCAGTTACTTGCACTGGTACCAGCA 119

Qy        115 GAAGCCAGGATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGT 174
              |||||||||||||||||||||||  |||||||||| ||||||||||||:|||||||||||
Db        120 GAAGCCAGGATCCTCCCCCAAACTCTGGATTTATAGCACATCCAACCTGGCTTCTGGAGT 179

Qy        175 CCCTGCTCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGT 234
              ||| ||||||||||||||| ||||||||||||||||||||||||||||| |||||||  |
Db        180 CCCAGCTCGCTTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGCAT 239

Qy        235 GGAGGCTGAAGATGCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTT 294
              ||||||||||||||||||||||||||||||||| ||||    ||||  ||| | ||||||
Db        240 GGAGGCTGAAGATGCTGCCACTTATTACTGCCACCAGTATCATCGTTCCCCGCTCACGTT 299

Qy        295 CGGAGGGGGGTCCAAGCTGGAAATAAAAGGTTCTA---CCTCTGGTTCTGGTAAATCTTC 351
              ||| |  ||| ||||||||||  | ||||     |   |||| || |||||       ||
Db        300 CGGTGCTGGGACCAAGCTGGAGCTCAAAGTCGACAAATCCTCAGGATCTGGCTCCGAATC 359

Qy        352 TGAAGGTAAAGGTGTGCAGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTC 411
                || | |     ||  |  ||  ||||||| ||| ||||| |||||| |||||||| |||
Db        360 CAAAAGCACGCAGGTCAAACTCGAGGAGTCTGGATCTGAGCTGGTGAGGCCTGGAGCTTC 419

Qy        412 AATGAAGATATCCTGCAAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGT 471
              | ||||| | ||||||||| ||||||| ||| |||||||  || ||   ||| |||||||
Db        420 AGTGAAGCTGTCCTGCAAGGCTTCTGGCTACACATTCACCAGCTACTGGATGCACTGGGT 479

Qy        472 GAAGCAGAGCCATGGAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGA 531
              ||||||||| |||||| |   ||||||||||||||||  |||| |||||   | ||||
Db        480 GAAGCAGAGGCATGGACAAGGCCTTGAGTGGATTGGAAATATTTATCCTGGTAGTGGTAG 539

Qy        532 TACTAACTACAACCAGAAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAG 591
              |||||||||| |  ||||||||||| ||||||||||| | ||||||||||  || |||||
Db        540 TACTAACTACGATGAGAAGTTCAAGAGCAAGGCCACACTGACTGTAGACACATCCTCCAG 599

Qy        592 CACAGCCTACATGGAGCTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGC 651
              ||||||||||||| | |||  ||| |||||||||||||||||||| |||||||||||| |
Db        600 CACAGCCTACATGCACCTCAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTAC 659

Qy        652 AAGGAGGGTTACGGACTGGTACTTCGATG---------TCTGGGGCGCAGGGACCACGGT 702
              ||| ||      | |||||  || | |||          ||||||   ||| ||| | ||
Db        660 AAGAAGCAGCCGGAACTGGGTCTACTATGCTATGGACTACTGGGGTCAAGGAACCTCAGT 719

Qy        703 CACCGTCTCC 712
              ||||||||||
Db        720 CACCGTCTCC 729


RESULT   6
Q55181
ID   Q55181 standard; cDNA; 720 BP.
AC   Q55181;
DT   21-JUL-1994  (first entry)
DE   Sequence encoding the single chain anti-erbB2 antibody, Ab no.21.
KW   Single chain anti-erbB1 antibody; cancer therapy; prevention;
KW   monoclonal antibody; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   cds             1. .171
FT                   /*tag=  a
PN   WO9400136-A.
PD   06-JAN-1994.
PF   21-OCT-1992; U08545.
PR   30-JUN-1992; US-906555.
PA   (MOLE-) MOLECULAR ONCOLOGY INC.
PI   Kasprzyk PG,  King CR;
DR   WPI; 94-025878/03.
DR   P-PSDB; R45443.
PT   Treatment of malignancies over-expressing ERB-[2 - using at least
PT   2 monoclonal antibodies which recognise different epitopes on
PT   gp185
PS   Example; Fig 8; 37pp; English.
CC   The source of human erbB-2 protein for the prodn. of antibodies no.
CC   23 and 21 (Q55180/R45442; Q55181/R45443) is a NIH/3T3 cell
CC   engineered to express the human erbB-2 protein on its surface (N/
CC   erbB-2). Abs no. 21 and 23 are directed against the extracellular
CC   domain of gp185 erbB-2. Nude mice manipulated to produce rapidly
CC   growing tumours were used in a trial of the efficacy of the Abs. In
CC   animals given a combination of the 2 Abs, tumours completely
CC   regressed after 11 days.
SQ   Sequence   720 BP;    184 A;    178 C;    179 G;    179 T;


  Query Match             59.3%;  Score 422;  DB 1;  Length 720;
  Best Local Similarity   79.3%;  Pred. No. 2.3e-103;
  Matches  511;  Conservative    1;  Mismatches  131;  Indels    1;  Gaps    1;

Qy          8 TGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTACACAATG 67
              |||||||||||||||||||||||||| ||||||||||||||||||| ||||| ||| |||
Db          2 TGCAGCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGT-CACCATG 60

Qy         68 ACTTGCAGGGCCACCCCAAGTGTAAGTTACATGCACTGGTATCAGCAGAAGCCAGGATCC 127
              || ||||| |||| | ||||||||||| ||||||||||||||||||||||| || |  ||
Db         61 ACCTGCAGTGCCAGCTCAAGTGTAAGTAACATGCACTGGTATCAGCAGAAGTCAAGCACC 120

Qy        128 TCCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGTCCCTGCTCGCTTC 187
              ||||||||||  ||| |||||  ||||||||| ||:|||||||||||||| | |||||||
Db        121 TCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTTC 180

Qy        188 AGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGAGGCTGAAGAT 247
```

```
          |||||| |||||||||| | |||||||||||||||  |||||||  ||||||||||||||
Db    181 AGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGAT 240

Qy    248 GCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGGAGGGGGGTCC 307
          |||||||||||||| ||  | ||| ||||| |   ||||  |||||||||   |||| |
Db    241 GCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGGACA 300

Qy    308 AAGCTGGAAATAAAAGGTTCTACCTCTGGTTCTGGTAAATCTTCTGAAGGTAAAGGTGTG 367
          ||| |||||||||||||||||||||| || ||||||||||||||||||||||||||||||
Db    301 AAGTTGGAAATAAAAGGTTCTACCTCCGGATCTGGTAAATCTTCTGAAGGTAAAGGTGTG 360

Qy    368 CAGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTCAATGAAGATATCCTGC 427
          ||||||||| |||| ||   |||| | ||       || |  | | | ||| | ||||||
Db    361 CAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGTCCTGC 420

Qy    428 AAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAAGCAGAGCCATGGA 487
          ||  ||||||  | |    | |  | | |    ||  |||||||||||||||| | || |
Db    421 AAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCCTGAA 480

Qy    488 AAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATACTAACTACAACCAG 547
           ||  ||| || |||||||||   ||| ||||| | |||||| |||||   ||  ||| |
Db    481 CAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATGACCCG 540

Qy    548 AAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCACAGCCTACATGGAG 607
          || ||| ||||||||||||   | || | |||||  || |||| | | |||||| |  ||
Db    541 AAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTACCTTCAG 600

Qy    608 CTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGC 651
          |||  ||| ||||||||||||||| |||| ||||||||||||||
Db    601 CTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGC 644
```

```
RESULT   7
T65007
ID   T65007 standard; cDNA; 720 BP.
AC   T65007;
DT   05-JUN-1997  (first entry)
DE   Single-chain anti-erbB2 antibody e21(Fv) cDNA.
KW   Single chain antibody; variable region; light chain; heavy chain;
KW   breast cancer; ovarian cancer; non-small cell lung carcinoma;
KW   immunodiagnosis; treatment; cytotoxic agent; erbB-2; ds.
OS   Mus musculus.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   mat_peptide     1. .720
FT                   /*tag= a
FT                   /product= e21(Fv)
PN   US5587458-A.
PD   24-DEC-1996.
PF   07-OCT-1991; 772270.
PR   07-OCT-1991; US-772270.
PR   30-JUN-1992; US-906555.
PR   14-MAY-1993; US-061092.
PA   (ARON-) ARONEX PHARM INC.
PI   Bird RE,  Kasprzyk PG,  King CR;
DR   WPI; 97-064831/06.
DR   P-PSDB; W15186.
PT   Single chain antibodies specific for erbB-2 protein, gp185  - with
PT   labels or cytotoxin, useful for detection and treatment of tumour
PT   cells expressing this protein
PS   Example 9; Columns 27-30; 28pp; English.
CC   The present cDNA sequence codes for a claimed single-chain antibody,
CC   designated e21(Fv), which binds to erbB-2. Monoclonal antibody e21
CC   was generated by immunising mice with N/erbB-2 cells overexpressing
CC   the gp185 protein, removing spleen cells and producing hybridomas
CC   by standard techniques. Messenger RNA coding for the anti-erbB-2
CC   monoclonal antibody was isolated and converted to cDNA. Regions
CC   coding for the heavy- and light- chain variable regions were then
CC   amplified by PCR and joined via a sequence encoding a peptide
CC   linker. The resulting single-chain antibody is useful for in vitro
CC   diagnosis of tumour cells which overexpress the erbB-2 gp185
CC   marker, e.g. breast, ovarian and non-small cell lung carcinomas,
CC   and, when coupled to a cytotoxic agent, to treat such tumours.
```

```
SQ   Sequence   720 BP;   184 A;   178 C;   179 G;   179 T;

  Query Match             59.3%;  Score 422;  DB 1;  Length 720;
  Best Local Similarity   79.3%;  Pred. No. 2.3e-103;
  Matches  511;  Conservative    1;  Mismatches  131;  Indels    1;  Gaps    1;

Qy      8 TGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTACACAATG 67
          |||||||||||||||||||||||||| ||||||||||||||||||| ||||| ||| |||
Db      2 TGCAGCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGT-CACCATG 60

Qy     68 ACTTGCAGGGCCACCCCAAGTGTAAGTTACATGCACTGGTATCAGCAGAAGCCAGGATCC 127
          || ||||| |||| | ||||||||||| ||||||||||||||||||||||| || |  ||
Db     61 ACCTGCAGTGCCAGCTCAAGTGTAAGTAACATGCACTGGTATCAGCAGAAGTCAAGCACC 120

Qy    128 TCCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGTCCCTGCTCGCTTC 187
          ||||||||||  ||| |||||  ||||||||| ||:|||||||||||||| | |||||||
Db    121 TCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTTC 180

Qy    188 AGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGAGGCTGAAGAT 247
          |||||| |||||||||| | |||||||||||||||  |||||||  ||||||||||||||
Db    181 AGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGAT 240

Qy    248 GCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGGAGGGGGGTCC 307
          |||||||||||||| ||  | ||| ||||| |   ||||  |||||||||   |||| |
Db    241 GCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGGACA 300

Qy    308 AAGCTGGAAATAAAAGGTTCTACCTCTGGTTCTGGTAAATCTTCTGAAGGTAAAGGTGTG 367
          ||| |||||||||||||||||||||| || ||||||||||||||||||||||||||||||
Db    301 AAGTTGGAAATAAAAGGTTCTACCTCCGGATCTGGTAAATCTTCTGAAGGTAAAGGTGTG 360

Qy    368 CAGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTCAATGAAGATATCCTGC 427
          ||||||||| |||| ||   |||| | ||       || |  | | | ||| | ||||||
Db    361 CAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGTCCTGC 420

Qy    428 AAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAAGCAGAGCCATGGA 487
          ||  ||||||  | |    | |  | | |    ||  |||||||||||||||| | || |
Db    421 AAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCCTGAA 480

Qy    488 AAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATACTAACTACAACCAG 547
           ||  ||| || |||||||||   ||| ||||| | |||||| |||||   ||  ||| |
Db    481 CAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATGACCCG 540

Qy    548 AAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCACAGCCTACATGGAG 607
          || ||| ||||||||||||   | || | |||||  || |||| | | |||||| |  ||
Db    541 AAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTACCTTCAG 600

Qy    608 CTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGC 651
          |||  ||| ||||||||||||||| |||| ||||||||||||||
Db    601 CTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGC 644
```

```
RESULT   8
Q81076
ID   Q81076 standard; cDNA; 913 BP.
AC   Q81076;
DT   14-MAR-1995  (first entry)
DE   Bispecific CD3-L6FvIg antibody derivative coding sequence.
KW   fusion protein; recombinant bispecific single chain antibody;
KW   helical peptide linker; anti-L6 antibody; tumour cell antigen;
KW   anti-CD3 antibody; variable region; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   signal_peptide  7. .75
FT                   /*tag= a
FT                   /note= "L6 light chain variable region leader"
FT   cds             7. .912
FT                   /*tag= b
FT   misc_rna        76. .819
FT                   /*tag= c
FT                   /note= "encodes CD3 VL-VH fusion"
FT   misc_rna        820. .828
FT                   /*tag= d
FT                   /note= "encodes hinge region of Fc domain"
FT   misc_rna        829. .912
FT                   /*tag= e
FT                   /note= "encodes helical peptide linker"
PN   EP-610046-A.
PD   10-AUG-1994.
PF   31-JAN-1994; 300692.
PR   01-FEB-1993; US-013420.
PR   13-SEP-1993; US-121054.
PA   (BRIM ) BRISTOL-MYERS SQUIBB CO.
PI   Bajorath J,  Fell PH,  Gilliland LK,  Hayden MS,  Ledbetter JA;
PI   Linsley PS;
DR   WPI; 94-250885/31.
DR   P-PSDB; R60206.
PT   Expression vector encoding bispecific fusion protein - having
PT   binding domains for separate targets joined by helical peptide,
PT   useful e.g. for diagnosis and treatment
PS   Example 1; Fig 11; 50pp; English.
CC   The VL and VH sequences of the anti-CD3 hybridoma G19-4 were
CC   amplified by PCR methods. A gene fusion was constructed from the
CC   two amplified domains and a (Gly4Ser)3 linker. The amino
CC   terminus of the VL-VH fusion cassette was fused at the SalI site to
CC   the L6 light chain variable region leader peptide and the
CC   carboxy-terminus was fused directly to the hinge region of the Fc
CC   domain at the BclI site and/or to a short "helical" peptide linker
CC   to construct the bispecific CD3-L6FvIg antibody derivative. The
CC   variable regions for L6 were fused in frame to the opposite end of
CC   the helical linker (not included in Q81076).
SQ   Sequence   913 BP;    252 A;    223 C;    228 G;    210 T;


  Query Match             50.0%;  Score 356.2;  DB 1;  Length 913;
  Best Local Similarity   71.9%;  Pred. No. 7.2e-86;
  Matches  537;  Conservative    0;  Mismatches  173;  Indels   37;  Gaps    4;

Qy         2 TGGACCTGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTAC 61
             | ||| | ||| |||| ||| || || |   ||||||||| ||||  || || |  || |
Db        74 TCGACATCCAGATGACACAGACTACATCCTCCCTGTCTGCCTCTCTGGGAGACAGAGT-C 132

Qy        62 ACAATGACTTGCAGGGCCACCC---CAAGTGTAAGTTACATGCACTGGTATCAGCAGAAG 118
             || || | ||||||||| |  |     | |   | |||  |  ||||||||||||||||
Db       133 ACCATCAGTTGCAGGGCAAGTCAGGACATTCGCAATTATTTAAACTGGTATCAGCAGAAA 192

Qy       119 CCAGGATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGTCCCT 178
             ||||      |    ||||    ||| ||   |||||| |   |    || ||||||||
Db       193 CCAGATGGAACTGTTAAACTCCTGATCTACTACACATCAAGATTACACTCAGGAGTCCCA 252

Qy       179 GCTCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGAG 238
              |  | ||||||||| |||||||||| ||   ||| ||||||||  |   ||   || |
Db       253 TCAAGGTTCAGTGGCAGTGGGTCTGGAACAGATTATTCTCTCACCATTGCCAACCTGCAA 312

Qy       239 GCTGAAGATGCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGGA 298
              | ||||||  ||||||||| |  ||||| ||| | | |      ||    ||||||||
Db       313 CCAGAAGATATTGCCACTTACTTTTGCCAACAGGGTAATACGCTTCCGTGGACGTTCGGT 372

Qy       299 GGGGGGTCCAAGCTGGAAATAAAAGGTTCTACCTCTGGTTCTGGTAAATCTTCTGAAGGT 358
             || ||  |||| |||| ||  ||| |      |  |||   |||         ||  ||
Db       373 GGAGGCACCAAACTGGTAACCAAACGGGAGCTCGGTGGCGGTGGCTCGGGCGGTGGTGGG 432

Qy       359 AAAGGT---------------------GTGCAGCTGCAGGAGTCAGGACCTGAGGTGGTG 397
                |||                     || ||||||||  |||| ||||||||  |||||
Db       433 TCGGGTGGCGGCGGATCTATCGATGAGGTCCAGCTGCAACAGTCTGGACCTGAACTGGTG 492

Qy       398 AAGCCTGGAGGTTCAATGAAGATATCCTGCAAGACTTCTGGTTACTCATTCACTGGCCAC 457
             |||||||||| ||||||||  || ||||||||| | ||||||||||||||||||||| ||
Db       493 AAGCCTGGAGCTTCAATGACAATGTCCTGCAAGGCCTCTGGTTACTCATTCACTGGCTAC 552

Qy       458 ACCATGAACTGGGTGAAGCAGAGCCATGGAAAGAACCTTGAGTGGATTGGACTTATTAAT 517
             | | |||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db       553 ATCGTGAACTGGCTGAAGCAGAGCCATGGAAAGAACCTTGAGTGGATTGGACTTATTAAT 612

Qy       518 CCTTACAATGGTGATACTAACTACAACCAGAAGTTCAAGGGCAAGGCCACATTTACTGTA 577
             || ||||| |||  ||||| |||||||||||| |||||||||||||||||||| ||||||
Db       613 CCATACAAAGGTCTTACTACCTACAACCAGAAATTCAAGGGCAAGGCCACATTAACTGTA 672

Qy       578 GACAAGTCGTCCAGCACAGCCTACATGGAGCTCCTCAGTCTGACATCTGAGGACTCTGCA 637
             |||||||| ||||||||||||||||||||||||||||||||||||||||| |||||||||
Db       673 GACAAGTCATCCAGCACAGCCTACATGGAGCTCCTCAGTCTGACATCTGAAGACTCTGCA 732

Qy       638 GTCTATTACTGTGCAAG------------GAGGGTTACGGACTGGTACTTCGATGTCTGG 685
             |||||||||||||||||              | |   |||||||||||||||||||||||
Db       733 GTCTATTACTGTGCAAGATCTGGGTACTATGGTGACTCGGACTGGTACTTCGATGTCTGG 792

Qy       686 GGCGCAGGGACCACGGTCACCGTCTCC 712
             |||||||||||||||  ||||||||||
Db       793 GGCGCAGGGACCACGTGCACCGTCTCC 819


RESULT   9
V00611
ID   V00611 standard; DNA; 906 BP.
AC   V00611;
DT   14-JUL-1998  (first entry)
DE   Anti-human SC single chain Fv/protamine fusion protein encoding DNA.
KW   Fusion protein; SC single chain Fv/protamine fusion protein; SECR;
KW   exogenous gene; serpin enzyme complex receptor; gene therapy; target;
KW   binding moiety; ds.
OS   Homo sapiens.
OS   Mus sp.
FH   Key             Location/Qualifiers
FT   CDS             1. .906
FT                   /*tag= a
FT                   /product= "anti-human SC single chain Fv/protamine
FT                   fusion protein"
PN   WO9746100-A1.
PD   11-DEC-1997.
PF   03-JUN-1997; U09858.
PR   03-JUN-1996; US-656906.
PA   (UYCA-) UNIV CASE WESTERN RESERVE.
PI   Davis PB, Ferkol TW, Ziady A;
DR   WPI; 98-041783/04.
DR   P-PSDB; W37085.
PT   Delivering compacted exogenous nucleic acid to cells  - by targeting
PT   the serpin enzyme complex receptor, used in gene therapy
PS   Example 9; Pages 120-121; 158pp; English.
CC   This DNA encodes an anti-human SC single chain Fv/protamine fusion
CC   protein containing a target binding moiety capable of binding to a
CC   serpin enzyme complex receptor (SECR), and a nucleic acid binding moiety.
CC   This can be used in a method for delivering an oligonucleotide to a
CC   mammalian cell. The method comprises conjugating the target binding
CC   moiety to a nucleic acid binding moiety to form a carrier and coupling
CC   the carrier to an expression vector encoding one or more gene products
CC   to form a pharmaceutical composition. A mammalian cell having on its
CC   surface SECR, is contacted with the pharmaceutical composition under
CC   conditions allowing binding to the receptor resulting in delivery of the
CC   pharmaceutical composition to the interior of the cell. The composition
CC   and method are used for the introduction of exogenous genetic material
CC   into target host cells expressing SECR on their surface. The nucleic acid
CC   may encode a functional wild-type or mutant gene or may be an antisense
CC   sequence or other nucleic acid having a therapeutic effect. The fusion
CC   protein may comprise a protein portion having therapeutic properties,
CC   e.g. enzymatic activity, cytokine activity and antibiotic activity which
CC   is delivered to a cell surface via the SECR binding moiety. The nucleic
CC   acid can be compacted at high concentrations with the carrier molecule at
CC   a critical salt concentration. The condensation of such complexes
CC   provides structural features to the DNA/cationic lipid complex that
CC   prolong in vivo expression.
SQ   Sequence   906 BP;    226 A;    231 C;    257 G;    192 T;


  Query Match             49.2%;  Score 350.4;  DB 1;  Length 906;
  Best Local Similarity   71.0%;  Pred. No. 2.5e-84;
  Matches  514;  Conservative    0;  Mismatches  191;  Indels   19;  Gaps    3;

Qy         1 ATGGACCTGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTA 60
             |||||| |   ||||||||||||||||||     || |||  |||| |||| ||| ||
Db       1 ATGGACATTGTGCTGACCCAGTCTCCAGCTTCTTTGGCTGTGTCTCTAGGGCAGAGGGC- 60

Qy      61 CACAATGACTTGCAGGGCCACCCCAAGTGTA---------------AGTTACATGCACTG 105
           ||| ||  | ||||| |||| |  ||||||                ||||  ||| ||||
Db      60 CACCATCTCCTGCAGAGCCAGCGAAAGTGTTGATAATTATGCCATTAGTTTTATGAACTG 119

Qy     106 GTATCAGCAGAAGCCAGGATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTKGC 165
           ||  || ||||| ||||||    | |||||||     || ||| |  |||||||||  |
Db     120 GTTCCAACAGAAACCAGGACAGCCACCCAAACTCCTCATCTATGCTGCATCCAACCAAGG 179

Qy     166 TTCTGGAGTCCCTGCTCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGT 225
            || || |||||||   | || |||||| |||||||||||||    | |   ||||   |
Db     180 ATCCGGGGTCCCTGGCAGGTTTAGTGGCAGTGGGTCTGGGACAGACTTCAGCCTCAACAT 239

Qy     226 CAGCAGAGTGGAGGCTGAAGATGCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCC 285
           |       ||||||  || ||| |||| |  |||| ||| |||||  | |        ||
Db     240 CCATCCTATGGAGGAGGATGATACTGCAATGTATTTCTGTCAGCAAAGTAAGGCGGTTCC 299

Qy     286 ACCCACGTTCGGAGGGGGGTCCAAGCTGGAAATAAA---AGGTTCTACCTCTGGTTCTGG 342
              |||||||||||||||  | |||||||| |||||   |||   |  ||| ||   |||
Db     300 GTACACGTTCGGAGGGGGCACAAAGCTGGAGATAAAAGGAGGCGGTGGCTCGGGCGGTGG 359

Qy     343 TAAATCTTCTGAAGGTAAAGGTGTGCAGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCC 402
              |||    |  ||    |  || ||||| ||| |||| ||||||||  ||||||||||
Db     360 CGGATCCGGCGGCGGCTCTGAGGTTCAGCTTCAGCAGTCTGGACCTGACCTGGTGAAGCC 419

Qy     403 TGGAGGTTCAATGAAGATATCCTGCAAGACTTCTGGTTACTCATTCACTGGCCACACCAT 462
           ||| | |||| ||||||||||||||||||||||||| ||| |||||| ||   | |||||
Db     420 TGGGGCTTCAGTGAAGATATCCTGCAAGACTTCTGGATACACATTCATTGAATATACCAT 479

Qy     463 GAACTGGGTGAAGCAGAGCCATGGAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTA 522
           | ||||||||||||||||||||||||||| ||||||||||||||||  |||||||||| |
Db     480 GCACTGGGTGAAGCAGAGCCATGGAAAGAGCCTTGAGTGGATTGGAGGTATTAATCCTAA 539

Qy     523 CAATGGTGATACTAACTACAACCAGAAGTTCAAGGGCAAGGCCACATTTACTGTAGACAA 582
           |||||||| |||||  |||||||||||||||||||||||||||||||| |||||||||||
Db     540 CAATGGTGGTACTAGTTACAACCAGAAGTTCAAGGGCAAGGCCACATTGACTGTAGACAA 599

Qy     583 GTCGTCCAGCACAGCCTACATGGAGCTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTA 642
           ||| |||| |||||||||||||||||||| | | |||||||||||||| |||||||||||
Db     600 GTCCTCCACCACAGCCTACATGGAGCTCCGCGGCCTGACATCTGAGGATTCTGCAGTCTA 659

Qy     643 TTACTGTGCAAGGAGGGTTACGGACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGT 702
           || |||||||||      || | ||       ||  ||    || |    ||  |   |
Db     660 TTCCTGTGCAAGATACTATAGGTACGACGTTCTCTCTGCTATGGACTACTGGGGCCAAGG 719

Qy     703 CACC 706
           ||||
Db     720 CACC 723


RESULT  10
Q73679
ID   Q73679 standard; DNA; 729 BP.
AC   Q73679;
DT   09-MAY-1995  (first entry)
DE   Fv(GP-4) immunosuppressive.
KW   MAb; monoclonal antibody; hybridoma; interleukin-2; IL-2;
KW   Fv; antibody variable region; GP-4; Fv(GP-4);
KW   immunosuppressive; ss.
OS   Mus sp.
PN   EP-621338-A.
PD   26-OCT-1994.
PF   21-APR-1994; 106257.
PR   21-APR-1993; JP-094491.
PR   07-MAR-1994; JP-036065.
PA   (AJIN ) AJINOMOTO KK.
PI   Hamura J,  Kanayama Y,  Nakazawa H,  Shimamura T,  Sugamura K;
PI   Takeshita T;
DR   WPI; 94-325948/41.
DR   P-PSDB; R60781.
PT   Immunosuppressant polypeptide - has ability to block
PT   interleukin-2 response
PS   Claim 18; Page 29; 37pp; English.
CC   MAb capable of binding to the gamma chain of the IL-2 receptor, and
CC   thus of blocking the IL-2 response, is produced by mouse hybridoma
CC   line GP-4 (FERM BP-4640).  DNA encoding the variable region of
CC   this MAb was expressed in E. coli, yielding Fv(GP-4) with
CC   immumosuppressive activity.
SQ   Sequence   729 BP;    199 A;    169 C;    182 G;    179 T;


  Query Match              48.9%;  Score 348.4;  DB 1;  Length 729;
  Best Local Similarity    71.5%;  Pred. No. 8e-84;
  Matches  519;  Conservative    0;  Mismatches  191;  Indels   16;  Gaps    4;

Qy       1 ATGGACCTGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTA 60
           |||||  | | |||||| |||||||||||   ||| |||||||||   || || |  ||
Db       1 ATGGATATTCTGCTGACACAGTCTCCAGCCTCCCTATCTGCATCTGTGGGAGAAACTGT- 60

Qy      61 CACAATGACTTGCAGGGCCACCCCAAGTGT---AAGTTACATGCACTGGTATCAGCAGAA 117
           ||| || || ||  | || |     | | |    | |||  |    ||||||||||||||
Db      60 CACCATCACATGTCGAGCAAGTGGGAATATTCACAATTATTTAGCATGGTATCAGCAGAA 119

Qy     118 GCCAGGATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGTCCC 177
            |  |||   || ||  | |    | | ||||   ||   | | | ||   ||| || ||
Db     120 ACAGGGAAAATCTCCTCAGCTCCTGGTCTATAATGCAAAAACCTTAGCAGATGGTGTGCC 179

Qy     178 TGCTCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGA 237
             |  | ||||||||| |||| || || ||    || |||||||   ||| |||  || |
Db     180 ATCAAGGTTCAGTGGCAGTGGATCAGGAACACAATATTCTCTCAAGATCAACAGCCTGCA 239

Qy     238 GGCTGAAGATGCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGG 297
           | ||||||||  ||  | ||||||||| || || |   |  |||  ||    ||||||||
Db     240 GCCTGAAGATTTTGGGAGTTATTACTGTCAACATTTTTGGAGTACTCCGTGGACGTTCGG 299

Qy     298 AGGGGGGTCCAAGCTGGAAATAAAAGGTTCTA---CCTCTGGTTCTGGTAAATCTTCTGA 354
            || ||| ||||||||||  | ||||     |   |||| || |||||       ||  |
Db     300 TGGAGGGACCAAGCTGGAGCTCAAAGTCGAGAAATCCTCAGGATCTGGCTCCGAATCCAA 359

Qy     355 AGGTAAAGGTGTGCAGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTCAAT 414
           | | |     ||  | ||  ||||||| ||||||||  ||||||||||||| | |||| |
Db     360 AAGCACGCAGGTCAAACTCGAGGAGTCTGGACCTGAGCTGGTGAAGCCTGGGGCTTCAGT 419

Qy     415 GAAGATATCCTGCAAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAA 474
           |||||||||||||||| ||||||||||||||||||||||| ||  |||| ||||||||||
Db     420 GAAGATATCCTGCAAGGCTTCTGGTTACTCATTCACTGGCTACTACATGCACTGGGTGAA 479

Qy     475 GCAGAGCCATGGAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATAC 534
           ||| ||||||| ||||| ||||||||||||||||| |||||||||||||||||||| |||
Db     480 GCAAAGCCATGTAAAGAGCCTTGAGTGGATTGGACGTATTAATCCTTACAATGGTGCTAC 539

Qy     535 TAACTACAACCAGAAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCAC 594
           || |||||||||||| ||||||| ||||||||  || |||||||| ||||| ||||||||
Db     540 TAGCTACAACCAGAATTTCAAGGACAAGGCCAGCTTGACTGTAGATAAGTCCTCCAGCAC 599

Qy     595 AGCCTACATGGAGCTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAG 654
           ||||||||||||||||| ||| ||||||||||||||||||||||||||||||||||||||
Db     600 AGCCTACATGGAGCTCCACAGCCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAG 659

Qy     655 GAGG---------GTTACGGACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCAC 705
              |         | ||    ||     || | |  |||||||  ||||||   ||||||
Db     660 AGAGAATTACTACGGTAGTAGCTACGGGTTTGCTTACTGGGGCCAAGGGACTCTGGTCAC 719

Qy     706 CGTCTC 711
            |||||
Db     720 TGTCTC 725


RESULT  11
Q41069
ID   Q41069 standard; DNA; 848 BP.
AC   Q41069;
DT   25-AUG-1993  (first entry)
DE   26-10 sFv coding sequence.
KW   Heavy; light; variable; VH; VL; region; antidigitoxin; monoclonal;
KW   antibody; MAb; 26-10; Fv; fragment; antigen; binding site; linker;
KW   expression-secretion system; T7 promoter; signal peptide; molecule;
KW   polymerase chain reaction; PCR; single chain; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   cds             6. .821
FT                   /*tag= a
FT   signal_peptide  6. .80
FT                   /*tag= b
FT                   /note= "PhoA signal sequence"
FT   misc_feature    81. .419
FT                   /*tag= c
FT                   /note= "26-10 VL coding sequence"
FT   misc_feature    420. .464
FT                   /*tag= d
FT                   /note= "Linker"
FT   misc_feature    465. .821
FT                   /*tag= e
FT                   /note= "26-10 VH coding sequence"
PN   WO9308300-A.
PD   29-APR-1993.
PF   16-OCT-1992; U08881.
PR   18-OCT-1991; US-777709.
PA   (UYCA-) UNIV CALGARY.
PI   Anthony JG, Ng SC, Wong S;
DR   WPI; 93-152491/18.
DR   P-PSDB; R34672.
PT   Expression-secretion vectors - for prodn. of biologically active
PT   antibody Fv fragments or single chain Fv molecules
PS   Disclosure; Fig 6; 51pp; English.
CC   This sequence encodes a single chain Fv molecule which contains
CC   the heavy and light variable portions (VH and VL) of antidigitoxin
CC   monoclonal antibody (MAb) 26-10, a signal sequence and a linker
CC   between the two variable regions. This sequence was used in the
CC   construction of an expression-secretion system for the production
CC   of biologically active Fv fragments. The system also contains a DNA
CC   sequence encoding the T7 promoter. The secretion-expression vector
CC   was produced by polymerase chain reaction (see also Q41070-81). Fv
CC   fragments comprising only the VH or VL regions, and single chain
CC   molecules may be used to specifically bind one or more of the same
CC   antigens as the full length antibody from which they are derived.
SQ   Sequence   848 BP;   209 A;   216 C;   219 G;   204 T;


  Query Match             46.6%;  Score 331.6;  DB 1;  Length 848;
  Best Local Similarity   72.9%;  Pred. No. 2.4e-79;
  Matches  459;  Conservative    0;  Mismatches  159;  Indels   12;  Gaps    2;

Qy      95 TACATGCACTGGTATCAGCAGAAGCCAGGATCCTCCCCCAAACCTTGGATTTATACCACA 154
           ||  |  ||||||| | ||||||| ||||    || || || |    ||| || |
Db     189 TATTTAAACTGGTACCTGCAGAAGGCAGGCCAGTCTCCAAAGCTCCTGATCTACAAAGTT 248

Qy     155 TCCAACCTKGCTTCTGGAGTCCCTGCTCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTAC 214
           |||||||    |||||| ||||| |   | ||||||||| |||| || |||||   || |
Db     249 TCCAACCGATTTTCTGGGGTCCCAGACAGGTTCAGTGGCAGTGGATCAGGGACAGATTTC 308

Qy     215 TCTCTCACAGTCAGCAGAGTGGAGGCTGAAGATGCTGCCACTTATTACTGCCAGCAGTGG 274
            | ||||   |||||||||||||||||||||||   |  | ||||| ||||   ||
Db     309 ACACTCAAGATCAGCAGAGTGGAGGCTGAAGATCTGGGAATTTATTTCTGCTCTCAAACT 368

Qy     275 AGTCGTAGCCCACCCACGTTCGGAGGGGGGTCCAAGCTGGAAATAAAAGGTTCTACCTCT 334
           |  | |   || || |||||||| || ||  ||||||||||||| ||| |   |   |||
Db     369 ACACATGTTCCTCCGACGTTCGGTGGAGGCACCAAGCTGGAAATCAAACGGGGTGGTTCT 428

Qy     335 GGTTCTGGTAAATCTTCTGAAGGTAAA---------GGTGTGCAGCTGCAGGAGTCAGGA 385
           ||||||||    ||   |   |||  |         |  || ||||||||  |||| |||
Db     429 GGTTCTGGCGGCTCCGGTTCCGGTGGATCCGGCTCTGAGGTCCAGCTGCAACAGTCCGGA 488

Qy     386 CCTGAGGTGGTGAAGCCTGGAGGTTCAATGAAGATATCCTGCAAGACTTCTGGTTACTCA 445
           |||||  ||||||||||||| | |||| ||| ||| ||||||||| ||||||| |||  |
Db     489 CCTGAACTGGTGAAGCCTGGGGCTTCAGTGAGGATGTCCTGCAAGTCTTCTGGATACATA 548

Qy     446 TTCACTGGCCACACCATGAACTGGGTGAAGCAGAGCCATGGAAAGAACCTTGAGTGGATT 505
           ||||||| |  |  |||||||||||||| ||||||||||||||||| |||||| |  |||
Db     549 TTCACTGACTTCTACATGAACTGGGTGAGGCAGAGCCATGGAAAGAGCCTTGATTACATT 608

Qy     506 GGACTTATTAATCCTTACAATGGTGATACTAACTACAACCAGAAGTTCAAGGGCAAGGCC 565
           |||  ||||  |||||||| ||||| ||||  ||||||||||||||||||||||||||||
Db     609 GGATATATTTCTCCTTACAGTGGTGTTACTGGCTACAACCAGAAGTTCAAGGGCAAGGCC 668

Qy     566 ACATTTACTGTAGACAAGTCGTCCAGCACAGCCTACATGGAGCTCCTCAGTCTGACATCT 625
           ||||| |||||||||||||| ||||||||||||||||||||||||| ||| ||||||||
Db     669 ACATTGACTGTAGACAAGTCCTCCAGCACAGCCTACATGGAGCTCCGCAGCCTGACATCG 728

Qy     626 GAGGACTCTGCAGTCTATTACTGTGCAAG---GAGGGTTACGGACTGGTACTTCGATGTC 682
           ||||| ||||||||||||||||||||| |   |  ||  |   | |||    | ||   |
Db     729 GAGGATTCTGCAGTCTATTACTGTGCAGGATCGTCGGGGAATAAGTGGGCTATGGACTAC 788

Qy     683 TGGGGCGCAGGGACCACGGTCACCGTCTCC 712
           |||||    ||  || | ||||||||||||
Db     789 TGGGGTCACGGAGCCTCAGTCACCGTCTCC 818


RESULT  12
Q12405
ID   Q12405 standard; DNA; 738 BP.
AC   Q12405;
DT   16-SEP-1991  (first entry)
DE   OVB3 light and heavy chains.
KW   Immunoglobulin; immunotoxin; antibody; variable region; PE40;
KW   pseudomonas exotoxin; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   cds             7. .729
FT                   /*tag= a
FT   misc_rna        7. .327
FT                   /*tag= b
FT                   /label= light chain
FT   misc_rna        328. .369
FT                   /*tag= b
FT                   /label= linker
FT   misc_rna        370. .729
FT                   /*tag= c
FT                   /label= heavy chain
PN   US7463111-A.
PD   28-MAY-1991.
PF   11-JAN-1990; 133066.
PR   11-JAN-1990; US-463111.
PA   (USSH ) NAT INST OF HEALTH.
PI   Pastan I;
DR   WPI; 91-200877/27.
DR   P-PSDB; R12798.
PT   Rapid cloning of antibody genes as single chain immuno-toxins -
PT   by fusing genes with DNA encoding cytotoxic molecules, to kill
PT   target cells
PS   Disclosure; Fig 3; 26pp; English.
CC   The sequence is an NdeI-HindIII fragment encoding the L and H
CC   chains of murine MAb OVB3, joined together via a linker. The se-
CC   quence is present in clone pOVB3158-1 in which the 3' end is
CC   linked to the 5' end of the coding sequence for PE40, a a Pseudo-
CC   monas exotoxin lacking domain I. The protein expressed by the
CC   clone can be used to kill specific target cells.
SQ   Sequence   738 BP;   181 A;   193 C;   192 G;   172 T;


  Query Match             45.0%;  Score 320.6;  DB 1;  Length 738;
  Best Local Similarity   67.6%;  Pred. No. 1.9e-76;
  Matches  479;  Conservative    1;  Mismatches  225;  Indels    4;  Gaps    2;

Qy       1 ATGGACCTGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTA 60
           |||||      ||| |||||||||||||||||| ||||||||||||||||||| |||||
Db       4 ATGGAAAATGTGCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGT- 63

Qy      61 CACAATGACTTGCAGGGCCACCCCAAGTGTAAGTTACATGCACTGGTATCAGCAGAAGCC 120
           ||| ||||| ||||| |||| | ||| ||||||||||||||||||||| ||||||||| |
Db      63 CACCATGACCTGCAGTGCCATCTCAACTGTAAGTTACATGCACTGGTACCAGCAGAAGTC 122
```

```
Qy     121 AGGATCCTCCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGTCCCTGC 180
           | |  ||||||||||||  |||||||||   |||||||| ||:|||||||||||||| |
Db     123 AAGCACCTCCCCCAAACTCTGGATTTATGAGACATCCAAACTGGCTTCTGGAGTCCCAGG 182

Qy     181 TCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGAGGC 240
           ||||||||||||| |||||||||| | |||||||||||||||  |||||||  |||||||
Db     183 TCGCTTCAGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGC 242

Qy     241 TGAAGATGCTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGGAGG 300
           |||||||| |||||||||||||||    ||| ||||| |   |||   ||||||||||||
Db     243 TGAAGATGTTGCCACTTATTACTGTTTTCAGGGGAGTGGGTACCCGTACACGTTCGGAGG 302

Qy     301 GGGGTCCAAGCTGGAAATAAAA---GGTTCTACCTCTGGTTCTGGTAAATCTTCTGAAGG 357
           |||| ||||||||||  | |||   |||    |||| || |||||       ||  |||
Db     303 GGGGACCAAGCTGGAGCTCAAAGAGGGTAAATCCTCAGGATCTGGCTCCGAATCCAAAGT 362

Qy     358 TAAAGGTGTGCAGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTCAATGAA 417
             | | ||||||||||  |||||| ||    |   | ||| |||||||||| ||   | |
Db     363 CGACGATGTGCAGCTGGTGGAGTCTGGGGGAGGCTTAGTGCAGCCTGGAGGGTCCCGGCA 422

Qy     418 GATATCCTGCAAGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAAGCA 477
             | |||||     | ||||| | | | |||| | ||      ||| |||||||    ||
Db     423 ACTCTCCTGTGCAGCCTCTGGATTCACTTTCAGTAGCTTTGGAATGCACTGGGTTCGTCA 482

Qy     478 GAGCCATGGAAAGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATACTAA 537
           |   |  |  |||   || |||||| | |     |||| || |     |  |   |   |
Db     483 GGCTCCAGAGAAGGGGCTGGAGTGGGTCGCATACATTAGTCGTGGCGGTAATACCATCTA 542

Qy     538 CTACAACCAGAAGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCACAGC 597
           |||     | |   | ||||||     |||  |  |   ||||||  |    | |||
Db     543 CTATGCAAACACAGTGAAGGGCCGATTCACCATCTCCAGAGACAATCCCAAGAACACCCT 602

Qy     598 CTACATGGAGCTCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGGAG 657
            | | || |  |   |||||| |  ||||| ||| | ||  | ||||||||||||||
Db     603 GTTCCTGCAAATGACCAGTCTAAGGTCTGACGACACGGCCATGTATTACTGTGCAAGATC 662

Qy     658 GGTTACGGACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACC 706
              |    || | |||||| |||    || |    ||     ||  |||
Db     663 CCATTACTACGGCTACTTCTATGCTATGGACTACTGGGGTCAGGGAACC 711

RESULT  13
T29058
ID   T29058 standard; DNA; 797 BP.
AC   T29058;
DT   03-OCT-1996  (first entry)
DE   CC49/212 SCA PLAP DNA construct.
KW   Single-chain antigen-binding fusion protein; SCA; immunoeffector;
KW   cytostatic; phospholipase A activating protein; PLAP; cancer;
KW   therapy; monoclonal antibody; CC49; ds.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   cds             1. .789
FT                   /*tag= a
FT                   /product= CC49 VL-212 linker-VH-PLAP fusion
PN   WO9611955-A1.
PD   25-APR-1996.
PF   13-OCT-1995; U12840.
PR   13-OCT-1994; US-323445.
PA   (ENZO-) ENZON INC.
PI   Filpula D,  Shorr R,  Whitlow M;
DR   WPI; 96-221949/22.
DR   P-PSDB; R97380.
PT   New antigen binding proteins contg. immunoeffector or cytolytic
PT   peptide - attached to variable regions of antibody light or heavy
PT   chain, useful in diagnosis and in destroying cancer cells
PS   Example 1; Fig 4A-B; 72pp; English.
CC   A novel DNA construct (T29058) codes for a single-chain antigen-
CC   binding fusion protein (R97380) comprising the light chain variable
CC   region (VL) of monoclonal antibody CC49 linked to the CC49 VH
CC   region via a spacer peptide (R97379). with a C-terminal
CC   phospholipase A activating protein (PLAP) immunoeffector peptide
CC   (R97368).  It was obtd. by ligating a synthetic PLAP fragment to an
CC   existing CC49/212 gene, and can be used to produce the fusion
CC   protein in transformed host cells.  The fusion protein combines
CC   the tumour antigen-binding capability of CC49 with the
CC   immunoeffector function of PLAP.
SQ   Sequence   797 BP;    196 A;    203 C;    195 G;    203 T;

  Query Match             44.8%;  Score 319;  DB 1;  Length 797;
  Best Local Similarity   72.0%;  Pred. No. 5.2e-76;
  Matches  447;  Conservative    1;  Mismatches  161;  Indels   12;  Gaps    2;

Qy      95 TACATGCACTGGTATCAGCAGAAGCCAGGATCCTCCCCCAAACCTTGGATTTATACCACA 154
           ||| ||  |||||| |||||||| |||||    || || ||||    ||||||     ||
Db     112 TACTTGGCCTGGTACCAGCAGAAACCAGGGCAGTCTCCTAAACTGCTGATTTACTGGGCA 171

Qy     155 TCCAACCTKGCTTCTGGAGTCCCTGCTCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTAC 214
           |||     :|  ||||| ||||||| ||||||||  ||| |||| ||||||||   || |
Db     172 TCCGCTAGGGAATCTGGGGTCCCTGATCGCTTCACAGGCAGTGGATCTGGGACAGATTTC 231

Qy     215 TCTCTCACAGTCAGCAGAGTGGAGGCTGAAGATGCTGCCACTTATTACTGCCAGCAGTGG 274
            ||||| |  ||||||| ||| || |||||||    ||   ||||||||| |||||||
Db     232 ACTCTCTCCATCAGCAGTGTGAAGACTGAAGACCTGGCAGTTTATTACTGTCAGCAGTAT 291

Qy     275 AGTCGTAGCCCACCCACGTTCGGAGGGGGGTCCAAGCTGGAAATAAAAGGTTCTACCTCT 334
             | |    || | ||||||||| |  ||| ||||||| |   | ||||| ||||| ||
Db     292 TATAGCTATCCCCTCACGTTCGGTGCTGGGACCAAGCTTGTGCTGAAAGGCTCTACTTCC 351

Qy     335 GGTTCTGGTAAATCTTCTGAAGGTAAAGGT---GTGCAGCTGCAGGAGTCAGGACCTGAG 391
           |||   || ||||| |||||||| ||||||   || ||||||||| |||| |   |||||
Db     352 GGTAGCGGCAAATCCTCTGAAGGCAAAGGTCAGGTTCAGCTGCAGCAGTCTGACGCTGAG 411

Qy     392 GTGGTGAAGCCTGGAGGTTCAATGAAGATATCCTGCAAGACTTCTGGTTACTCATTCACT 451
            ||||||| ||||| | |||| ||||||| ||||||||| ||||||| ||| | ||||||
Db     412 TTGGTGAAACCTGGGGCTTCAGTGAAGATTTCCTGCAAGGCTTCTGGCTACACCTTCACT 471

Qy     452 GGCCACACCATGAACTGGGTGAAGCAGAGCCATGGAAAGAACCTTGAGTGGATTGGACTT 511
           | |||  | ||   ||||||||| |||| || || | ||  ||| || |||||||||  |
Db     472 GACCATGCAATTCACTGGGTGAAACAGAACCCTGAACAGGGCCTGGAATGGATTGGATAT 531

Qy     512 ATTAATCCTTACAATGGTGATACTAACTACAACCAGAAGTTCAAGGGCAAGGCCACATTT 571
            ||  |||    |||| ||||  ||| |||||  ||| ||||||||||||||||||| |
Db     532 TTTTCTCCCGGAAATGATGATTTTAAATACAATGAGAGGTTCAAGGGCAAGGCCACACTG 591

Qy     572 ACTGTAGACAAGTCGTCCAGCACAGCCTACATGGAGCTCCTCAGTCTGACATCTGAGGAC 631
           |||| |||||| || |||||||| |||||| || |||||  ||| ||||||||||||||
Db     592 ACTGCAGACAAATCCTCCAGCACTGCCTACGTGCAGCTCAACAGCCTGACATCTGAGGAT 651

Qy     632 TCTGCAGTCTATTACTGTGCAAGGAGGGTTACGGACTGGTACTTCGATGTCTGGGGCGCA 691
           |||||||| |||| |||| ||||            |  | |  | |    ||||||   |
Db     652 TCTGCAGTGTATTTCTGTACAAG---------ATCCCTGAATATGGCCTACTGGGGTCAA 702

Qy     692 GGGACCACGGTCACCGTCTCC 712
           || ||| ||||||||||||||
Db     703 GGAACCTCGGTCACCGTCTCC 723

RESULT  14
Q43291
ID   Q43291 standard; DNA; 1460 BP.
AC   Q43291;
DT   27-SEP-1993  (first entry)
DE   Sequence encoding divalent CC49/212 single-chain antigen-binding
DE   protein.
KW   Antibody; multivalent; variable region; heavy chain; light chain;
KW   linker; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   cds             1. .1458
FT                   /*tag= a
FT   cds             1. .339
FT                   /*tag= b
FT                   /product= CC49 VL
FT   cds             340. .374
FT                   /*tag=  c
FT                   /product= 212 linker
FT   cds             375. .426
FT                   /*tag=  d
FT                   /product= CC49 VH
FT   cds             427. .1065
FT                   /*tag=  e
FT                   /product= CC49 VL
FT   cds             1066. .1107
FT                   /*tag=  f
FT                   /product= 212 linker
FT   cds             1108. .1449
FT                   /*tag=  g
FT                   /product= CC49 VH
PN   WO9311161-A.
PD   10-JUN-1993.
PF   20-NOV-1992; U09965.
PR   25-NOV-1991; US-796936.
PA   (ENZO-) ENZON INC.
PI   Bird RE,  Filpula D,  Hardman K,  Rollence M,  Whitlow MD,
PI   Wood JF;
DR   WPI; 93-196999/24.
DR   P-PSDB; R37649.
PT   New multivalent antigen-binding protein e.g. contg. CC49 Mab
PT   regions - useful in diagnosis, for destroying blood clots and
PT   targetting cytotoxic agents or enzymes to tumour cells
PS   Example; Fig 16A; 118pp; English.
CC   The number 212 refers to a 14-residue linker. CC49 is the CC49
CC   antibody.
SQ   Sequence   1460 BP;    363 A;    364 C;    365 G;    368 T;


  Query Match             44.6%;  Score 317.8;  DB 1;  Length 1460;
  Best Local Similarity   73.9%;  Pred. No. 1.3e-75;
  Matches  416;  Conservative    1;  Mismatches  143;  Indels    3;  Gaps    1;

Qy     95 TACATGCACTGGTATCAGCAGAAGCCAGGATCCTCCCCCAAACCTTGGATTTATACCACA 154
          ||| ||  |||||| |||||||| |||||    || || ||||    ||||||     ||
Db    112 TACTTGGCCTGGTACCAGCAGAAACCAGGGCAGTCTCCTAAACTGCTGATTTACTGGGCA 171

Qy    155 TCCAACCTKGCTTCTGGAGTCCCTGCTCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTAC 214
          |||     :|  ||||| ||||||| ||||||||  ||| |||| ||||||||   || |
Db    172 TCCGCTAGGGAATCTGGGGTCCCTGATCGCTTCACAGGCAGTGGATCTGGGACAGATTTC 231

Qy    215 TCTCTCACAGTCAGCAGAGTGGAGGCTGAAGATGCTGCCACTTATTACTGCCAGCAGTGG 274
           ||||| |  ||||||| ||| || |||||||    ||   ||||||||| |||||||
Db    232 ACTCTCTCCATCAGCAGTGTGAAGACTGAAGACCTGGCAGTTTATTACTGTCAGCAGTAT 291

Qy    275 AGTCGTAGCCCACCCACGTTCGGAGGGGGGTCCAAGCTGGAAATAAAAGGTTCTACCTCT 334
            | |    || | ||||||||| |  ||| ||||||| |   | ||||| ||||| ||
Db    292 TATAGCTATCCCCTCACGTTCGGTGCTGGGACCAAGCTTGTGCTGAAAGGCTCTACTTCC 351

Qy    335 GGTTCTGGTAAATCTTCTGAAGGTAAAGGT---GTGCAGCTGCAGGAGTCAGGACCTGAG 391
          |||   || ||||| |||||||| ||||||   || ||||||||| |||| |   |||||
Db    352 GGTAGCGGCAAATCCTCTGAAGGCAAAGGTCAGGTTCAGCTGCAGCAGTCTGACGCTGAG 411

Qy    392 GTGGTGAAGCCTGGAGGTTCAATGAAGATATCCTGCAAGACTTCTGGTTACTCATTCACT 451
           ||||||| ||||| | |||| ||||||| ||||||||| ||||||| ||| | ||||||
Db    412 TTGGTGAAACCTGGGGCTTCAGTGAAGATTTCCTGCAAGGCTTCTGGCTACACCTTCACT 471

Qy    452 GGCCACACCATGAACTGGGTGAAGCAGAGCCATGGAAAGAACCTTGAGTGGATTGGACTT 511
          | |||  | ||  |||||||||| |||| || || | ||  ||| || |||||||||  |
Db    472 GACCATGCAATTCACTGGGTGAAACAGAACCCTGAACAGGGCCTGGAATGGATTGGATAT 531

Qy    512 ATTAATCCTTACAATGGTGATACTAACTACAACCAGAAGTTCAAGGGCAAGGCCACATTT 571
           ||  |||    |||| ||||  ||| |||||  ||| ||||||||||||||||||| |
Db    532 TTTTCTCCCGGAAATGATGATTTTAAATACAATGAGAGGTTCAAGGGCAAGGCCACACTG 591

Qy    572 ACTGTAGACAAGTCGTCCAGCACAGCCTACATGGAGCTCCTCAGTCTGACATCTGAGGAC 631
          |||| |||||| || |||||||| |||||| || |||||  ||| ||||||||||||||
Db    592 ACTGCAGACAAATCCTCCAGCACTGCCTACGTGCAGCTCAACAGCCTGACATCTGAGGAT 651

Qy    632 TCTGCAGTCTATTACTGTGCAAG 654
          |||||||| |||| |||| ||||
Db    652 TCTGCAGTGTATTTCTGTACAAG 674


RESULT  15
T29059
ID   T29059 standard; DNA; 803 BP.
AC   T29059;
DT   03-OCT-1996  (first entry)
DE   PLAP CC49/212 SCA DNA construct.
KW   Single chain antigen-binding fusion protein; SCA; immunoeffector;
KW   cytostatic; phospholipase A activating protein; PLAP; cancer;
KW   therapy; monoclonal antibody; CC49; ds.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   cds             1. .795
FT                   /*tag=  a
FT                   /product= PLAP CC49 VL-212 linker-VH fusion
PN   WO9611955-A1.
PD   25-APR-1996.
PF   13-OCT-1995; U12840.
PR   13-OCT-1994; US-323445.
PA   (ENZO-) ENZON INC.
PI   Filpula D,  Shorr R,  Whitlow M;
DR   WPI; 96-221949/22.
DR   P-PSDB; R97381.
PT   New antigen binding proteins contg. immunoeffector or cytolytic
PT   peptide - attached to variable regions of antibody light or heavy
PT   chain, useful in diagnosis and in destroying cancer cells
PS   Disclosure; Page 39-40; 72pp; English.
CC   A novel DNA construct (T29059) codes for a single-chain antigen
CC   binding fusion protein (R97381) comprising the light chain variable
CC   region (VL) of monoclonal antibody CC49 linked to the CC49 VH
CC   region via a spacer peptide (R97379) and with an N-terminal
CC   phospholipase A activating protein (PLAP) immunoeffector peptide
CC   (R97368).  It can be used to produce the fusion protein in
CC   transformed host cells.  The fusion protein combines the tumour
CC   antigen-binding capability of CC49 with the immunoeffector function
CC   of PLAP.
SQ   Sequence   803 BP;    198 A;    205 C;    196 G;    204 T;


  Query Match             44.6%;  Score 317.8;  DB 1;  Length 803;
  Best Local Similarity   73.9%;  Pred. No. 1.1e-75;
  Matches  416;  Conservative    1;  Mismatches  143;  Indels    3;  Gaps    1;

Qy     95 TACATGCACTGGTATCAGCAGAAGCCAGGATCCTCCCCCAAACCTTGGATTTATACCACA 154
          ||| ||  |||||| |||||||| |||||    || || ||||    ||||||     ||
Db    181 TACTTGGCCTGGTACCAGCAGAAACCAGGGCAGTCTCCTAAACTGCTGATTTACTGGGCA 240

Qy    155 TCCAACCTKGCTTCTGGAGTCCCTGCTCGCTTCAGTGGCGGTGGGTCTGGGACCTCTTAC 214
          |||     :|  ||||| ||||||| ||||||||  ||| |||| ||||||||   || |
Db    241 TCCGCTAGGGAATCTGGGGTCCCTGATCGCTTCACAGGCAGTGGATCTGGGACAGATTTC 300

Qy    215 TCTCTCACAGTCAGCAGAGTGGAGGCTGAAGATGCTGCCACTTATTACTGCCAGCAGTGG 274
           ||||| |  ||||||| ||| || |||||||    ||   ||||||||| |||||||
Db    301 ACTCTCTCCATCAGCAGTGTGAAGACTGAAGACCTGGCAGTTTATTACTGTCAGCAGTAT 360

Qy    275 AGTCGTAGCCCACCCACGTTCGGAGGGGGGTCCAAGCTGGAAATAAAAGGTTCTACCTCT 334
            | |    || | ||||||||| |  ||| ||||||| |   | ||||| ||||| ||
Db    361 TATAGCTATCCCCTCACGTTCGGTGCTGGGACCAAGCTTGTGCTGAAAGGCTCTACTTCC 420

Qy    335 GGTTCTGGTAAATCTTCTGAAGGTAAAGGT---GTGCAGCTGCAGGAGTCAGGACCTGAG 391
          |||   || ||||| |||||||| ||||||   || ||||||||| |||| |   |||||
Db    421 GGTAGCGGCAAATCCTCTGAAGGCAAAGGTCAGGTTCAGCTGCAGCAGTCTGACGCTGAG 480

Qy    392 GTGGTGAAGCCTGGAGGTTCAATGAAGATATCCTGCAAGACTTCTGGTTACTCATTCACT 451
           ||||||| ||||| | |||| ||||||| ||||||||| ||||||| ||| | ||||||
Db    481 TTGGTGAAACCTGGGGCTTCAGTGAAGATTTCCTGCAAGGCTTCTGGCTACACCTTCACT 540

Qy    452 GGCCACACCATGAACTGGGTGAAGCAGAGCCATGGAAAGAACCTTGAGTGGATTGGACTT 511
          | |||  | ||  |||||||||| |||| || || | ||  ||| || |||||||||  |
Db    541 GACCATGCAATTCACTGGGTGAAACAGAACCCTGAACAGGGCCTGGAATGGATTGGATAT 600

Qy    512 ATTAATCCTTACAATGGTGATACTAACTACAACCAGAAGTTCAAGGGCAAGGCCACATTT 571
           ||  |||    |||| ||||  ||| |||||  ||| ||||||||||||||||||| |
Db    601 TTTTCTCCCGGAAATGATGATTTTAAATACAATGAGAGGTTCAAGGGCAAGGCCACACTG 660

Qy    572 ACTGTAGACAAGTCGTCCAGCACAGCCTACATGGAGCTCCTCAGTCTGACATCTGAGGAC 631
          |||| |||||| || |||||||| |||||| || |||||  ||| ||||||||||||||
Db    661 ACTGCAGACAAATCCTCCAGCACTGCCTACGTGCAGCTCAACAGCCTGACATCTGAGGAT 720

Qy    632 TCTGCAGTCTATTACTGTGCAAG 654
          |||||||| |||| |||| ||||
Db    721 TCTGCAGTGTATTTCTGTACAAG 743


Search completed: May 13, 1999, 09:55:24
Job time: 1182 sec
```

```
                    GenCore version 4.5
         Copyright (c) 1993 - 1998  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         May 13, 1999, 09:54:13 ; Search time 736.46 Seconds
                                         (without alignments)
                                         3497.589 Million cell updates/sec

Title:          US-08-704-178-2
Perfect score:  720
Sequence:       1 ATGCAGCTGACCCAGTCTCC..........CGGTCACCGTCTCCTCATAA 720

Scoring table:  IDENTITY_NUC

Searched:       808301 seqs, 1788773984 residues

Database :      GenEmbl:*
                1:  gb_bal:*
                2:  gb_ba2:*
                3:  gb_in:*
                4:  gb_om:*
                5:  gb_ov:*
                6:  gb_pat:*
                7:  gb_ph:*
                8:  gb_pl1:*
                9:  gb_pl2:*
                10:  gb_pr1:*
                11:  gb_pr2:*
                12:  gb_pr3:*
                13:  gb_ro:*
                14:  gb_st:*
                15:  gb_sy:*
                16:  gb_un:*
                17:  gb_vi:*
                18:  gb_htg:*
                19:  em_ba:*
                20:  em_fun:*
                21:  em_hum1:*
                22:  em_hum2:*
                23:  em_in:*
                24:  em_om:*
                25:  em_or:*
                26:  em_ov:*
                27:  em_pat:*
                28:  em_ph:*
                29:  em_pl:*
                30:  em_ro:*
                31:  em_sy:*
                32:  em_un:*
                33:  em_vi:*
                34:  em_htg:*
                35:  em_sts:*
                36:  gb_bal:*
                37:  gb_ba2:*
                38:  gb_pl1:*
                39:  gb_pl2:*
                40:  gb_pr1:*
                41:  gb_pr2:*
                42:  gb_pr3:*
                43:  gb_sts:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                 SUMMARIES
                 %
Result           Query
  No.    Score   Match Length DB  ID                     Description
------------------------------------------------------------------------
    1      720  100.0    720  6  I32407                I32407 Sequence 2
    2    433.4   60.2    711  6  I32406                I32406 Sequence 1
    3    433.4   60.2    711 15  S39590                 S39590 anti-erbB2
    4      432   60.0    732  6  I31035                I31035 Sequence 1
    5      432   60.0    732 27  E10361                 E10361 cDNA encodi
    6    404.4   56.2    738 15  SYNAR3IGA             M34000 Synthetic m
    7    383.8   53.3    720  6  E13599                E13599 DNA encodin
    8    333.8   46.4    828 13  MMZ95478              Z95478 Mus musculu
    9    327.6   45.5    822 13  MMZ95477              Z95477 Mus musculu
   10    326.4   45.3    735  6  E13598                E13598 DNA encodin
   11    321.2   44.6    741  6  I08678                I08678 Sequence 10
   12    311.6   43.3    753  6  I08679                I08679 Sequence 12
   13      299   41.5    729  6  I31036                I31036 Sequence 3
   14      299   41.5    729 27  E10362                 E10362 cDNA encodi
   15    297.4   41.3    744 15  SYNDA44X               L17037 Synthetic s
   16    293.6   40.8    803  6  AR012824              AR012824 Sequence
   17    292.6   40.6    797  6  AR012822              AR012822 Sequence
   18      289   40.1    819  6  A18692                A18692 Synthetic n
   19    288.8   40.1    861 13  MMCEAHCH               X52769 M.musculus
   20    286.8   39.8    711  6  I08677                I08677 Sequence 8
   21    283.4   39.4    401 13  MUSIGLDH              M64623 Mouse rearr
   22    282.8   39.3    444 13  MUSIGHDZ              M13328 Mouse Ig mu
   23    282.8   39.3    444 13  MUSIGHEA              M13329 Mouse Ig mu
   24    282.4   39.2    417 13  MUSIGHNPH             M19570 Mouse Ig ac
   25    280.8   39.0    511 13  MUSIGHBC              K00608 Mouse Ig ac
   26    280.2   38.9    318 13  AF057545              AF057545 Mus muscu
   27      278   38.6    321 13  MUSIGLAU              M20464 Mouse (clon
   28    277.4   38.5    321 13  MUSIGLAV              M20465 Mouse (clon
   29    273.4   38.0    299 13  MMU55616              U55616 Mus musculu
   30    273.2   37.9    321 13  MUSIGLAX              M20467 Mouse (clon
   31    272.6   37.9    321 13  AF021870              AF021870 Mus muscu
   32    271.8   37.8    723 13  AF035617              AF035617 Mus muscu
   33    270.4   37.6    369 13  MMVRHC                X93024 M.musculus
   34      270   37.5    501 13  MUSIGHBH              K00687 Mouse Ig ac
   35      270   37.5    501 13  MMIGVK1               X00894 Mouse mRNA
   36    269.6   37.4    390 13  MUSIGHBB              M13832 Mouse Ig ac
   37    269.6   37.4    304 13  MMU55603              U55603 Mus musculu
   38    269.6   37.4    315 13  MMIG1KL2              X02551 Mouse mRNA
   39    269.6   37.4    420 13  MMIGGNP2              X02563 Mouse mRNA
   40    269.6   37.4    354 13  MDIGMVAP              Z22078 M.domesticu
   41    269.2   37.4    321 13  MUSIGLAW              M20466 Mouse (clon
   42    269.2   37.4    456 13  MMIGHI7H              X14624 Murine MRL-
   43    268.4   37.3    351 13  MUSIGHAVAC            L22322 Mouse rearr
   44    268.4   37.3    414 13  MUSIGHNPF             M12749 Mouse Ig ac
   45    268.4   37.3    414 13  MMIGGNP7              X02568 Mouse mRNA

                                 ALIGNMENTS

RESULT   1
I32407
LOCUS       I32407         720 bp    DNA             PAT       07-JAN-1997
DEFINITION  Sequence 2 from patent US 5587458.
ACCESSION   I32407
NID         g1823198
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 720)
  AUTHORS   King,C.Richter, Kasprzyk,P.G. and Bird,R.E.
  TITLE     Anti-erbB-2 antibodies, combinations thereof, and therapeutic and
            diagnostic uses thereof
  JOURNAL   Patent: US 5587458-A 2 24-DEC-1996;
FEATURES             Location/Qualifiers
     source          1. .720
                     /organism="unknown"
BASE COUNT      184 a    178 c    179 g    179 t
ORIGIN

  Query Match             100.0%;  Score 720;  DB 6;  Length 720;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-211;
  Matches  720;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 ATGCAGCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATG 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 ATGCAGCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATG 60

Qy       61 ACCTGCAGTGCCAGCTCAAGTGTAAGTAACATGCACTGGTATCAGCAGAAGTCAAGCACC 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 ACCTGCAGTGCCAGCTCAAGTGTAAGTAACATGCACTGGTATCAGCAGAAGTCAAGCACC 120

Qy      121 TCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTTC 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 TCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTTC 180

Qy      181 AGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGAT 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 AGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGAT 240

Qy      241 GCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGGACA 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 GCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGGACA 300

Qy      301 AAGTTGGAAATAAAAGGTTCTACCTCCGGATCTGGTAAATCTTCTGAAGGTAAAGGTGTG 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 AAGTTGGAAATAAAAGGTTCTACCTCCGGATCTGGTAAATCTTCTGAAGGTAAAGGTGTG 360

Qy      361 CAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGTCCTGC 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 CAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGTCCTGC 420

Qy      421 AAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCCTGAA 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 AAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCCTGAA 480

Qy      481 CAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATGACCCG 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 CAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATGACCCG 540

Qy      541 AAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTACCTTCAG 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 AAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTACCTTCAG 600

Qy      601 CTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGCTTCTTATTACTACTAT 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 CTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGCTTCTTATTACTACTAT 660

Qy      661 AGTGCTTACTATGCTATGTACTACTGGGGTCAAGGAACCTCGGTCACCGTCTCCTCATAA 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661 AGTGCTTACTATGCTATGTACTACTGGGGTCAAGGAACCTCGGTCACCGTCTCCTCATAA 720


RESULT    2
I32406
LOCUS       I32406        711 bp    DNA             PAT       07-JAN-1997
DEFINITION  Sequence 1 from patent US 5587458.
ACCESSION   I32406
NID         g1823197
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 711)
  AUTHORS   King,C.Richter, Kasprzyk,P.G. and Bird,R.E.
  TITLE     Anti-erbB-2 antibodies, combinations thereof, and therapeutic and
            diagnostic uses thereof
  JOURNAL   Patent: US 5587458-A 1 24-DEC-1996;
FEATURES             Location/Qualifiers
     source          1. .711
                     /organism="unknown"
BASE COUNT      175 a    182 c    190 g    164 t
ORIGIN
```

```
  Query Match             60.2%;  Score 433.4;  DB 6;  Length 711;
  Best Local Similarity   79.6%;  Pred. No. 3.4e-123;
  Matches  512;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy        2 TGCAGCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATGA 61
            |||||||||||||||||||||||||| |||||||||||||||||||| |||||||| ||||
Db        8 TGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTCACAATGA 67

Qy       62 CCTGCAGTGCCAGCTCAAGTGTAAGTAACATGCACTGGTATCAGCAGAAGTCAAGCACCT 121
            | ||||| |||| | |||||||||||| ||||||||||||||||||||||| || |  |||
Db       68 CTTGCAGGGCCACCCCAAGTGTAAGTTACATGCACTGGTATCAGCAGAAGCCAGGATCCT 127

Qy      122 CCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTTCA 181
            |||||||||   ||| |||||  |||||||||| ||||||||||||||||||| | ||||||||
Db      128 CCCCCAAACCTTGGATTTATACCACATCCAACCTGGCTTCTGGAGTCCCTGCTCGCTTCA 187

Qy      182 GTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGATG 241
            ||||| |||||||||| | ||||||||||||||||  |||||||  ||||||||||||||||
Db      188 GTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGAGGCTGAAGATG 247

Qy      242 CTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGGACAA 301
            ||||||||||||| ||  | ||| ||||| |   ||||  |||||||||   |||| | |
Db      248 CTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGGAGGGGGGTCCA 307

Qy      302 AGTTGGAAATAAAAGGTTCTACCTCCGGATCTGGTAAATCTTCTGAAGGTAAAGGTGTGC 361
            || ||||||||||||||||||||||| || ||||||||||||||||||||||||||||||||
Db      308 AGCTGGAAATAAAAGGTTCTACCTCTGGTTCTGGTAAATCTTCTGAAGGTAAAGGTGTGC 367

Qy      362 AGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGTCCTGCA 421
            |||||||| |||| ||   |||| | ||       || |  | | | ||| | |||||||
Db      368 AGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTCAATGAAGATATCCTGCA 427

Qy      422 AAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCCTGAAC 481
            |  ||||||  | |    | |  | | |    ||  |||||||||||||||| | || | 
Db      428 AGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAAGCAGAGCCATGGAA 487

Qy      482 AGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATGACCCGA 541
            ||  ||| || |||||||||   ||| ||||| | |||||| |||||   ||  ||| ||
Db      488 AGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATACTAACTACAACCAGA 547

Qy      542 AATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTACCTTCAGC 601
            | ||| ||||||||||||   | || | |||||  || |||| | | |||||| |  |||
Db      548 AGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCACAGCCTACATGGAGC 607

Qy      602 TCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGC 644
            ||  ||| ||||||||||||||||| |||| ||||||||||||||
Db      608 TCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGC 650


RESULT    3
S39590
LOCUS       S39590        711 bp                    SYN       10-FEB-1993
DEFINITION  anti-erbB2 imnmunotoxin antigen binding region [mice, Other
            Synthetic Recombinant Partial, 711 nt].
ACCESSION   S39590
NID         g251113
KEYWORDS    .
SOURCE      Mus sp.
  ORGANISM  Mus sp.
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 711)
  AUTHORS   Batra,J.K., Kasprzyk,P.G., Bird,R.E., Pastan,I. and King,C.R.
  TITLE     Recombinant anti-erbB2 immunotoxins containing Pseudomonas exotoxin
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 89 (13), 5867-5871 (1992)
  MEDLINE   92335198
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 108547] from the original journal article.
            This sequence comes from Fig. 1.
FEATURES             Location/Qualifiers
     source          1. .711
```

```
                     /organism="Mus sp."
                     /db_xref="taxon:10095"
     gene            1. .711
                     /partial
                     /gene="anti-erbB2 imnmunotoxin antigen binding region"
     CDS             1. .711
                     /partial
                     /gene="anti-erbB2 imnmunotoxin antigen binding region"
                     /note="This sequence comes from Fig. 1."
                     /codon_start=1
                     /db_xref="PID:g251114"
                     /translation="MDLQLTQSPAILSASPGEKVTMTCRATPSVSYMHWYQQKPGSSP
                     KPWIYTTSNLASGVPARFSGGGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGS
                     KLEIKGSTSGSGKSSEGKGVQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQS
                     HGKNLEWIGLINPYNGDTNYNQKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCAR
                     RVTDWYFDVWGAGTTVTVS"
BASE COUNT      175 a    182 c    190 g    164 t
ORIGIN


  Query Match             60.2%;  Score 433.4;  DB 15;  Length 711;
  Best Local Similarity   79.6%;  Pred. No. 3.4e-123;
  Matches  512;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy        2 TGCAGCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATGA 61
            |||||||||||||||||||||||||| ||||||||||||||||||| |||||||| ||||
Db        8 TGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTCACAATGA 67

Qy       62 CCTGCAGTGCCAGCTCAAGTGTAAGTAACATGCACTGGTATCAGCAGAAGTCAAGCACCT 121
            | ||||| |||| | ||||||||||| ||||||||||||||||||||||| || |  |||
Db       68 CTTGCAGGGCCACCCCAAGTGTAAGTTACATGCACTGGTATCAGCAGAAGCCAGGATCCT 127

Qy      122 CCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTTCA 181
            |||||||||  ||| |||||  ||||||||| ||||||||||||||||| | ||||||||
Db      128 CCCCCAAACCTTGGATTTATACCACATCCAACCTGGCTTCTGGAGTCCCTGCTCGCTTCA 187

Qy      182 GTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGATG 241
            ||||| |||||||||| | |||||||||||||||  |||||||  |||||||||||||||
Db      188 GTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGAGGCTGAAGATG 247

Qy      242 CTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGGACAA 301
            ||||||||||||| ||  | ||| ||||| |   ||||  |||||||||   |||| | |
Db      248 CTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGGAGGGGGGTCCA 307

Qy      302 AGTTGGAAATAAAAGGTTCTACCTCCGGATCTGGTAAATCTTCTGAAGGTAAAGGTGTGC 361
            || |||||||||||||||||||||| || |||||||||||||||||||||||||||||||
Db      308 AGCTGGAAATAAAAGGTTCTACCTCTGGTTCTGGTAAATCTTCTGAAGGTAAAGGTGTGC 367

Qy      362 AGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGTCCTGCA 421
            |||||||| |||| ||   |||| | ||       || |  | | | ||| | |||||||
Db      368 AGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTCAATGAAGATATCCTGCA 427

Qy      422 AAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCCTGAAC 481
            |  ||||||  | |    | |  | | |    ||  |||||||||||||||| | || | 
Db      428 AGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAAGCAGAGCCATGGAA 487

Qy      482 AGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATGACCCGA 541
            ||  ||| || |||||||||   ||| ||||| | |||||| |||||   ||  ||| ||
Db      488 AGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATACTAACTACAACCAGA 547

Qy      542 AATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTACCTTCAGC 601
            | ||| ||||||||||||   | || | |||||  || |||| | | |||||| |  |||
Db      548 AGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCACAGCCTACATGGAGC 607

Qy      602 TCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGC 644
            ||  ||| ||||||||||||||| |||| ||||||||||||||
Db      608 TCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGC 650


RESULT    4
I31035
LOCUS       I31035        732 bp    DNA             PAT       20-DEC-1996
DEFINITION  Sequence 1 from patent US 5582826.
```

```
ACCESSION   I31035
NID         g1821826
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 732)
  AUTHORS   Shimamura,T., Hamuro,J., Nakazawa,H., Kanayama,Y., Sugamura,K. and
            Takeshita,T.
  TITLE     Monoclonal antibodies which bind the gamma chain of human
            interleukin-2 receptor
  JOURNAL   Patent: US 5582826-A 1 10-DEC-1996;
FEATURES             Location/Qualifiers
     source          1. .732
                     /organism="unknown"
BASE COUNT      180 a    204 c    182 g    166 t
ORIGIN


  Query Match             60.0%;  Score 432;  DB 6;  Length 732;
  Best Local Similarity   76.7%;  Pred. No. 9.3e-123;
  Matches  557;  Conservative    0;  Mismatches  160;  Indels    9;  Gaps    2;

Qy        1 ATGCAGCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATG 60
            || | |||||||||||||||| |||||||||||||||||| |||||||  ||||||||||
Db        7 ATCCTGCTGACCCAGTCTCCATCAATCATGTCTGCATCTCTAGGGGAACGGGTCACCATG 66

Qy       61 ACCTGCAGTGCCAGCTCAAGTGTA------AGTAACATGCACTGGTATCAGCAGAAGTCA 114
            ||||||| ||||||||||||||||      ||| || |||||||||| ||||||||| ||
Db       67 ACCTGCACTGCCAGCTCAAGTGTAAGTTCCAGTTACTTGCACTGGTACCAGCAGAAGCCA 126

Qy      115 AGCACCTCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGT 174
             |  ||||||||||||||||| |||||  ||||||||| ||||||||||||||||||| |
Db      127 GGATCCTCCCCCAAACTCTGGATTTATAGCACATCCAACCTGGCTTCTGGAGTCCCAGCT 186

Qy      175 CGCTTCAGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCT 234
            ||||||||||||||||||||||| | ||||||||||||||| ||||||||||||||||||
Db      187 CGCTTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGCATGGAGGCT 246

Qy      235 GAAGATGCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCG 294
            |||||||||||||||||||| ||  | |||     | | | |||  ||||||||||  | 
Db      247 GAAGATGCTGCCACTTATTACTGCCACCAGTATCATCGTTCCCCGCTCACGTTCGGTGCT 306

Qy      295 GGGACAAAGTTGGAAATAAAAGGTTCTA---CCTCCGGATCTGGTAAATCTTCTGAAGGT 351
            ||||| ||| ||||  | ||||     |   |||| ||||||||       ||  || | 
Db      307 GGGACCAAGCTGGAGCTCAAAGTCGACAAATCCTCAGGATCTGGCTCCGAATCCAAAAGC 366

Qy      352 AAAGGTGTGCAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAG 411
            |     ||  | ||  || |||||||   |||||| ||  |    || || | ||| |||
Db      367 ACGCAGGTCAAACTCGAGGAGTCTGGATCTGAGCTGGTGAGGCCTGGAGCTTCAGTGAAG 426

Qy      412 TTGTCCTGCAAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAG 471
             |||||||||| ||||||| || ||   | |    ||| |  || |||||||||||||||
Db      427 CTGTCCTGCAAGGCTTCTGGCTACACATTCACCAGCTACTGGATGCACTGGGTGAAGCAG 486

Qy      472 AGGCCTGAACAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTA 531
            |||| || ||| ||||| || |||||||||   ||| ||||||  | ||||| ||||   
Db      487 AGGCATGGACAAGGCCTTGAGTGGATTGGAAATATTTATCCTGGTAGTGGTAGTACTAAC 546

Qy      532 TATGACCCGAAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCC 591
            || ||   ||| ||| || |||||||||   | || | |||||||||||||| | | |||
Db      547 TACGATGAGAAGTTCAAGAGCAAGGCCACACTGACTGTAGACACATCCTCCAGCACAGCC 606

Qy      592 TACCTTCAGCTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGCTTCTTAT 651
            ||| | || |||||||||||||||||||||||| |||| |||||||||||| |
Db      607 TACATGCACCTCAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTACAAGAAGC 666

Qy      652 TACTACTATAGTGCTTACTATGCTATGTACTACTGGGGTCAAGGAACCTCGGTCACCGTC 711
              |    |  | |  |||||||||||| |||||||||||||||||||||| |||||||||
Db      667 AGCCGGAACTGGGTCTACTATGCTATGGACTACTGGGGTCAAGGAACCTCAGTCACCGTC 726

Qy      712 TCCTCA 717
            ||||||
Db    727 TCCTCA 732


RESULT   5
E10361
ID   E10361     standard; RNA; ROD; 732 BP.
XX
AC   E10361;
XX
NI   d1108698
XX
DT   08-OCT-1997 (Rel. 52, Created)
DT   08-OCT-1997 (Rel. 52, Last updated, Version 1)
XX
DE   cDNA encoding a monoclonal antibody against human Interleukin 2
DE   receptor gamma chain.
XX
KW   JP 1995313188-A/1.
XX
OS   Mus sp.
OC   Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Rodentia;
OC   Sciurognathi; Muridae; Murinae; Mus.
XX
RN   [1]
RP   1-732
RA   Shimamura T., Hamuro J., Nakazawa H., Kanayama Y., Sugamura K.,
RA   Takeshita T.;
RT   "IMMUNOSUPPRESSIVE AGENT";
RL   Patent number JP 1995313188-A/1, 05-DEC-1995.
RL   AJINOMOTO CO INC, SUGAMURA KAZUO,.
XX
CC   OS   Mus sp. (mouse)
CC   PN   JP 1995313188-A/1
CC   PD   05-DEC-1995
CC   PF   21-APR-1994 JP 1994082836
CC   PR   21-APR-1993 JP 93P    94491, 07-MAR-1994 JP 94P    36065
CC   PI   SHIMAMURA TOSHIAKI, HAMURO JUNJI, NAKAZAWA HARUMI,
CC   PI   KANAYAMA YUKA,
CC   PI   SUGAMURA KAZUO, TAKESHITA TOSHIICHI
CC   PC   C12P21/08,A61K39/395,A61K39/395,C12N1/21,C12N5/20,
CC   PC   C12N15/13//C12N15/06,
CC   PC   (C12P21/08,C12R1:19),(C12P21/08,C12R1:91),(C12N1/21,C12R1:19);
CC   CC   strandedness: Double;
CC   CC   topology: Linear;
CC   FH   Key             Location/Qualifiers
CC   FH
CC   FT   source          1. .732
CC   FT                   /organism="Mus sp."
CC   FT                   /cell_type="hybridoma"
CC   FT                   /cell_line="GP-2"
CC   FT   mat_peptide     1. .732
CC   FT                   /product="anti-IL-2 receptor gamma chain"
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .732
FT                   /organism="Mus sp."
XX
SQ   Sequence 732 BP; 180 A; 204 C; 182 G; 166 T; 0 other;


  Query Match             60.0%;  Score 432;  DB 27;  Length 732;
  Best Local Similarity   76.7%;  Pred. No. 9.3e-123;
  Matches  557;  Conservative    0;  Mismatches  160;  Indels    9;  Gaps    2;

Qy      1 ATGCAGCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATG 60
          || | ||||||||||||||| |||||||||||||||||||| |||||||  ||||||||||
Db      7 ATCCTGCTGACCCAGTCTCCATCAATCATGTCTGCATCTCTAGGGGAACGGTCACCATG 66

Qy     61 ACCTGCAGTGCCAGCTCAAGTGTA------AGTAACATGCACTGGTATCAGCAGAAGTCA 114
          ||||||| ||||||||||||||||      ||| || |||||||||| |||||||||| ||
Db     67 ACCTGCACTGCCAGCTCAAGTGTAAGTTCCAGTTACTTGCACTGGTACCAGCAGAAGCCA 126
```

```
Qy    115 AGCACCTCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGT 174
           |  |||||||||||||||| |||||  ||||||||| |||||||||||||||||||| |
Db    127 GGATCCTCCCCCAAACTCTGGATTTATAGCACATCCAACCTGGCTTCTGGAGTCCCAGCT 186

Qy    175 CGCTTCAGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCT 234
          ||||||||||||||||||||||| | |||||||||||||||| ||||||||||||||||||
Db    187 CGCTTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGCATGGAGGCT 246

Qy    235 GAAGATGCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCG 294
          |||||||||||||||||||| ||  | |||     | | | |||  ||||||||||  | 
Db    247 GAAGATGCTGCCACTTATTACTGCCACCAGTATCATCGTTCCCCGCTCACGTTCGGTGCT 306

Qy    295 GGGACAAAGTTGGAAATAAAAGGTTCTA---CCTCCGGATCTGGTAAATCTTCTGAAGGT 351
          ||||| ||| ||||  | ||||     |   |||| ||||||||       ||  || | 
Db    307 GGGACCAAGCTGGAGCTCAAAGTCGACAAATCCTCAGGATCTGGCTCCGAATCCAAAAGC 366

Qy    352 AAAGGTGTGCAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAG 411
          |     ||  | ||  || |||||||   ||||||| ||  |    || || | ||| |||
Db    367 ACGCAGGTCAAACTCGAGGAGTCTGGATCTGAGCTGGTGAGGCCTGGAGCTTCAGTGAAG 426

Qy    412 TTGTCCTGCAAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAG 471
           ||||||||||| ||||||| || ||    | |    ||| |  || ||||||||||||||||
Db    427 CTGTCCTGCAAGGCTTCTGGCTACACATTCACCAGCTACTGGATGCACTGGGTGAAGCAG 486

Qy    472 AGGCCTGAACAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTA 531
          |||| || ||| ||||| || |||||||||   ||| ||||||  | ||||| ||||   
Db    487 AGGCATGGACAAGGCCTTGAGTGGATTGGAAATATTTATCCTGGTAGTGGTAGTACTAAC 546

Qy    532 TATGACCCGAAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCC 591
          || ||   ||| ||| || |||||||||   | || | |||||||||||||| | | |||
Db    547 TACGATGAGAAGTTCAAGAGCAAGGCCACACTGACTGTAGACACATCCTCCAGCACAGCC 606

Qy    592 TACCTTCAGCTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGCTTCTTAT 651
          ||| | |  |||||||||||||||||||||||||| |||| ||||||||||||| |       
Db    607 TACATGCACCTCAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTACAAGAAGC 666

Qy    652 TACTACTATAGTGCTTACTATGCTATGTACTACTGGGGTCAAGGAACCTCGGTCACCGTC 711
            |    |  | |  ||||||||||||| |||||||||||||||||||||| |||||||||
Db    667 AGCCGGAACTGGGTCTACTATGCTATGGACTACTGGGGTCAAGGAACCTCAGTCACCGTC 726

Qy    712 TCCTCA 717
          ||||||
Db    727 TCCTCA 732


RESULT   6
SYNAR3IGA
LOCUS       SYNAR3IGA     738 bp    DNA             SYN       15-SEP-1990
DEFINITION  Synthetic mouse Ig heavy and light chain antibody (OVB3), partial
            cds.
ACCESSION   M34000
NID         g207974
KEYWORDS    C-region.
SOURCE      Mouse hybridoma, cDNA to mRNA, and E.coli vector DNA.
  ORGANISM  artificial sequence
            artificial sequence.
REFERENCE   1  (bases 1 to 738)
  AUTHORS   Chaudhary,V.K., Batra,J.K., Gallo,M.G., Willingham,M.C.,
            Fitzgerald,D.J. and Pastan,I.
  TITLE     A rapid method of cloning functional variable-region antibody genes
            in Escherichia coli as single-chain immunotoxins
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 87, 1066-1070 (1990)
  MEDLINE   90138938
FEATURES             Location/Qualifiers
     source          1. .738
                     /organism="artificial sequence"
                     /db_xref="taxon:29278"
     CDS             <1. .>738
                     /note="OVB3 Ig light and heavy chains"
                     /codon_start=1
                     /transl_table=11
                     /db_xref="PID:g207975"
                     /translation="HMENVLTQSPAIMSASPGEKVTMTCSAISTVSYMHWYQQKSSTS
```

```
                   PKLWIYETSKLASGVPGRFSGSGSGNSYSLTISSMEAEDVATYYCFQGSGYPYTFGGG
                   TKLELKEGKSSGSGSESKVDDVQLVESGGGLVQPGGSRQLSCAASGFTFSSFGMHWVR
                   QAPEKGLEWVAYISRGGNTIYYANTVKGRFTISRDNPKNTLFLQMTSLRSDDTAMYYC
                   ARSHYYGYFYAMDYWGQGTTLTVSSKAF"
     misc_feature  326. .365
                   /note="linker DNA"
BASE COUNT      181 a    193 c    192 g    172 t
ORIGIN

  Query Match             56.2%;  Score 404.4;  DB 15;  Length 738;
  Best Local Similarity   73.7%;  Pred. No. 3.1e-114;
  Matches  529;  Conservative    0;  Mismatches  186;  Indels    3;  Gaps    1;

Qy       6 GCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATGACCTG 65
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15 GCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATGACCTG 74

Qy      66 CAGTGCCAGCTCAAGTGTAAGTAACATGCACTGGTATCAGCAGAAGTCAAGCACCTCCCC 125
           |||||||| ||||| ||||||| ||||||||||||| |||||||||||||||||||||||
Db      75 CAGTGCCATCTCAACTGTAAGTTACATGCACTGGTACCAGCAGAAGTCAAGCACCTCCCC 134

Qy     126 CAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTTCAGTGG 185
           |||||||||| ||||||| |||||||||||||||||||||||||||||||||||||||||
Db     135 CAAACTCTGGATTTATGAGACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTTCAGTGG 194

Qy     186 CAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGATGCTGC 245
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     195 CAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGATGTTGC 254

Qy     246 CACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGGACAAAGTT 305
           ||||||||| |||| |||||||||||||||||||| | ||||||||   |||||| ||| |
Db     255 CACTTATTACTGTTTTCAGGGGAGTGGGTACCCGTACACGTTCGGAGGGGGGACCAAGCT 314

Qy     306 GGAAATAAAA---GGTTCTACCTCCGGATCTGGTAAATCTTCTGAAGGTAAAGGTGTGCA 362
           |||  | |||   |||    |||| ||||||||       ||  |||   | | ||||||
Db     315 GGAGCTCAAAGAGGGTAAATCCTCAGGATCTGGCTCCGAATCCAAAGTCGACGATGTGCA 374

Qy     363 GCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGTCCTGCAA 422
           ||||  | |||||||||  |   | || |     || |  |      |  | |||||
Db     375 GCTGGTGGAGTCTGGGGGAGGCTTAGTGCAGCCTGGAGGGTCCCGGCAACTCTCCTGTGC 434

Qy     423 AGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCCTGAACA 482
           ||| ||||  ||||   | |    || |   || ||||||||    |||   || ||  |
Db     435 AGCCTCTGGATTCACTTTCAGTAGCTTTGGAATGCACTGGGTTCGTCAGGCTCCAGAGAA 494

Qy     483 GGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATGACCCGAA 542
           ||| ||||| ||| | | ||  |||  ||||    |  ||  |     ||||     |
Db     495 GGGGCTGGAGTGGGTCGCATACATTAGTCGTGGCGGTAATACCATCTACTATGCAAACAC 554

Qy     543 ATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTACCTTCAGCT 602
           | |  |||||     ||  ||  |   |||||   ||   ||| |    | ||| ||  |
Db     555 AGTGAAGGGCCGATTCACCATCTCCAGAGACAATCCCAAGAACACCCTGTTCCTGCAAAT 614

Qy     603 CAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGCTTCTTATTACTACTATAG 662
            | ||| || |  ||||| ||||| ||| | ||||||||||    |   | |||||  |
Db     615 GACCAGTCTAAGGTCTGACGACACGGCCATGTATTACTGTGCAAGATCCCATTACTACGG 674

Qy     663 TGCTTACTATGCTATGTACTACTGGGGTCAAGGAACCTCGGTCACCGTCTCCTCATAA 720
               | ||||||||||| ||||||||||||| |||||| |  |||| |||||||||| ||
Db     675 CTACTTCTATGCTATGGACTACTGGGGTCAGGGAACCACTCTCACAGTCTCCTCAAAA 732

RESULT    7
E13599
LOCUS       E13599        720 bp    DNA             PAT       27-APR-1998
DEFINITION  DNA encoding a signal chain Fv antibody of anti-gp130 antibody
            GPX7.
ACCESSION   E13599
NID         g3252401
KEYWORDS    JP 1997220092-A/2.
SOURCE      unidentified.
  ORGANISM  unidentified
```

```
            unclassified.
REFERENCE   1  (bases 1 to 720)
  AUTHORS   Ekida,T., Yasukawa,K., Imanaka,T. and Takagi,M. .
  TITLE     PRODUCTION OF SINGLE-STRAND FV ANTIBODY
  JOURNAL   Patent: JP 1997220092-A 2 26-AUG-1997;
            TOSOH CORP
COMMENT     OS   None
            OC   Artificial sequences.
            PN   JP 1997220092-A/2
            PD   26-AUG-1997
            PF   15-FEB-1996 JP 1996027622
            PI   EKIDA TEIJI, YASUKAWA KIYOSHI, IMANAKA TADAYUKI, PI   TAKAGI
            MASAHIRO
            PC   C12N15/09,C12N1/21,C12P21/08,(C12N1/21,C12R1:19),(C12P21/08,
            PC   C12R1:19);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .720
            FT                   /organism='Artificial sequences'.
FEATURES             Location/Qualifiers
     source          1. .720
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      169 a    193 c    190 g    168 t
ORIGIN

  Query Match             53.3%;  Score 383.8;  DB 6;  Length 720;
  Best Local Similarity   73.0%;  Pred. No. 6.9e-108;
  Matches  522;  Conservative    0;  Mismatches  187;  Indels    6;  Gaps    2;

Qy       6 GCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATGACCTG 65
           ||| ||||||||||||||||||| | || |||||||||||||||||| |||||||||||||||||
Db       9 GCTCACCCAGTCTCCAGCATTGATATCTGCATCTCCAGGGGAGAAGGTCACCATGACCTG 68

Qy      66 CAGTGCCAGCTCAAGTGTAAGTAACATGCACTGGTATCAGCAGAAGTCAAGCACCTCCCC 125
           || || ||||||||||||| |  |||| | ||||| |||||||||| || |  |||||||
Db      69 CAATGTCAGCTCAAGTGTTACTTCCATGTATTGGTACCAGCAGAAGCCAGGATCCTCCCC 128

Qy     126 CAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTTCAGTGG 185
           ||||| |||| |||||  ||||||||| |||||||||||||||||   ||||| ||||||
Db     129 CAAACCCTGGATTTATCTCACATCCAACCTGGCTTCTGGAGTCCCTCCTCGCTCCAGTGG 188

Qy     186 CAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGATGCTGC 245
           |||||||||||| | | ||||||||||||| ||||||||| |||||||||||||||||||
Db     189 CAGTGGGTCTGGGACCACTTACTCTCTCACAATCAGCAGCTTGGAGGCTGAAGATGCTGC 248

Qy     246 CACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGGACAAAGTT 305
           ||||||||| ||  | ||| |||||    ||||  ||||||||||  | ||||| ||| |
Db     249 CACTTATTACTGCCAGCAGTGGAGTACTAACCCGCTCACGTTCGGTGCTGGGACCAAGCT 308

Qy     306 GGAAATAAAAGGTTCTACCTCCGGATCTGGTAAATCTTCTGAAGGTAAA---GGTGTGCA 362
            ||  |    ||   |  ||| ||   ||||   ||   ||  ||   |   |  || ||
Db     309 CGAGCTGGGTGGCGGTGGCTCGGGCGGTGGTGGGTCGGGTGGCGGCGGATCGGACGTCCA 368

Qy     363 GCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGTCCTGCAA 422
             |||||||||||||   ||| || ||       ||||| | ||| |||   |  |||||||
Db     369 ATTGCAGCAGTCTGGACCTGAACTGGTGAAGCCTGGGGCTTCAGTGAAGATACCCTGCAA 428

Qy     423 AGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCCTGAACA 482
            ||||| |  | ||   | |  |||||  | ||  |||||||||||||||| | || | |
Db     429 GGCTTCAGGATACACATTCACTGACTACAACATGGACTGGGTGAAGCAGAGCCATGGAAA 488

Qy     483 GGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATGACCCGAA 542
           | |||| || |||||||||   ||| || || | | ||||  || | | ||  |||  ||
Db     489 GAGCCTTGAGTGGATTGGAGATATTAATTCTCATAGTGGTGGTATTATCTACAACCAAAA 548

Qy     543 ATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTACCTTCAGCT 602
            ||| ||| |||||||   | || | ||| |  |||||| | | | |||||| |  ||||
Db     549 GTTCAAGGACAAGGCCATATTGACTGTAGATAAGTCCTCCAGCACAGCCTACATGGAGCT 608

Qy     603 CAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGCTTCTTATTACTACTATAG 662
           | ||||||||||||||||||||||||| |||||||| ||||   |      |||||||
Db     609 CCGCAGCCTGACATCTGAGGACACTGCAGTCTATTATTGTG---CAAGAACCTACTATGT 665

Qy     663 TGCTTACTATGCTATGTACTACTGGGGTCAAGGAACCTCGGTCACCGTCTCCTCA 717
                 | |     |  |||||||||| ||||| ||| |  |||| |||||||||
Db     666 GTACGGCCACTTCTTTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCA 720


RESULT   8
MMZ95478
LOCUS       MMZ95478       828 bp    RNA             ROD       29-OCT-1997
DEFINITION  Mus musculus mRNA for IgG3/kappa antibody, scFv49-CK.
ACCESSION   Z95478 AF004405
NID         g2108312
KEYWORDS    IgG3.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 828)
  AUTHORS   Schouten,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-1997) Dept. of Nematology, Wageningen
            Agricultural University, P.O. Box 8123, Wageningen 6700 ES, The
            Netherlands
REFERENCE   2  (bases 1 to 828)
  AUTHORS   Schouten,A., Roosien,J., de Boer,J.M., Wilmink,A., Rosso,M.N.,
            Bosch,D., Stiekema,W.J., Gommers,F.J., Bakker,J. and Schots,A.
  TITLE     Improving scFv antibody expression levels in the plant cytosol
  JOURNAL   FEBS Lett. 415 (2), 235-241 (1997)
  MEDLINE   98010486
FEATURES             Location/Qualifiers
     source          1. .828
                     /organism="Mus musculus"
                     /note="recombinant gene"
                     /db_xref="taxon:10090"
     CDS             1. .828
                     /codon_start=1
                     /product="IgG3/kappa antibody"
                     /db_xref="PID:e1167936"
                     /db_xref="PID:g2108313"
                     /translation="MAQVDGDIQMTQSPPSLPVSLGDQASISCRSSQSIVHSKGNTYL
                     EWFLQKPGQSPKLLIYKVSNRFSGVPDRFSGSGSGTDFTLKISRLEAEDLGVYYCFQG
                     SHVPYTFGGGTKLEIKGGGGSGGGGSGGGGSEVQLQESGAELVRSGASVKLSCTASDF
                     NIKDYYIHWVKQRPEQGLEWIGWIDPENGGTDYAPKFQGKATMTADTSSNTAYLQLSS
                     LTSEDTAVYYCNAYGNYDGYRGQGTTVTVSSAKTTVGAAAEQKLISEEDLNDIKDEL"
     misc_feature    19. .354
                     /note="Light chain variable domain (Vl) coding sequence"
     misc_feature    355. .399
                     /note="Linker peptide coding sequence"
     misc_feature    400. .747
                     /note="Heavy chain variable domain (Vh) coding sequence"
     misc_feature    748. .762
                     /note="Heavy chain first constant domain (CH1) fragment
                     coding sequence"
     misc_feature    775. .807
                     /note="c-myc tag coding sequence"
     misc_feature    814. .825
                     /note="KDEL coding sequence"
BASE COUNT      218 a    196 c    228 g    186 t
ORIGIN


  Query Match             46.4%;  Score 333.8;  DB 13;  Length 828;
  Best Local Similarity   70.3%;  Pred. No. 1.9e-92;
  Matches  487;  Conservative    0;  Mismatches  182;  Indels   24;  Gaps    2;

Qy       1 ATGCAGCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATG 60
           || ||| |||| ||||||||| |   | || |||    ||  || ||  | | | ||||
Db      22 ATCCAGATGACACAGTCTCCACCCTCCCTGCCTGTCAGTCTTGGAGATCAAGCCTCCATC 81

Qy      61 ACCTGCAGTGCCAGCTCAAGTGT------------------AAGTAACATGCACTGGTAT 102
            | |||||  | ||    ||  |                   |   |  |  | ||||
Db      82 TCTTGCAGATCTAGTCAGAGCATTGTACATAGTAAAGGAAACACCTATTTAGAATGGTTC 141

Qy     103 CAGCAGAAGTCAAGCACCTCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCT 162
           | ||||||  || ||   || || || |||  | | ||  |    ||||| |    ||||
Db     142 CTGCAGAAACCAGGCCAGTCTCCAAAGCTCCTGATCTACAAAGTTTCCAACCGATTTTCT 201

Qy     163 GGAGTCCCAGGTCGCTTCAGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGC 222
           || |||||||   | |||||||||||||| || || |    || | | |||| |||||||
Db     202 GGGGTCCCAGACAGGTTCAGTGGCAGTGGATCAGGGACAGATTTCACACTCAAGATCAGC 261

Qy     223 AGCATGGAGGCTGAAGATGCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTC 282
           ||  |||||||||| |||   |    |||||| || | ||| ||          || | |
Db     262 AGACTGGAGGCTGAGGATCTGGGAGTTTATTACTGCTTTCAAGGTTCACATGTTCCGTAC 321

Qy     283 ACGTTCGGCTCGGGGACAAAGTTGGAAATAAAA------GGTTCTACCTCCGGATCTGGT 336
           ||||||||   ||||||||||||||||||||||      |||  |   |  ||    ||
Db     322 ACGTTCGGAGGGGGGACAAAGTTGGAAATAAAAGGGGGTGGTGGTTCTGGAGGTGGAGGA 381

Qy     337 AAATCTTCTGAAGGTAAAGGTGTGCAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGA 396
                |   | ||||   |  || ||||||||  |||||||||  ||||||||  |   |
Db     382 TCCGGTGGAGGAGGTTCTGAGGTCCAGCTGCAAGAGTCTGGGGCAGAGCTTGTGAGGTCA 441

Qy     397 GGGGCCTTAGTCAAGTTGTCCTGCAAAGCTTCTGACTTCAACATTAAAGACTATTATATC 456
           ||||||| ||||||||||||||||| ||||||||||||||||||||||||||| |||||
Db     442 GGGGCCTCAGTCAAGTTGTCCTGCACAGCTTCTGACTTCAACATTAAAGACTACTATATA 501

Qy     457 CACTGGGTGAAGCAGAGGCCTGAACAGGGCCTGGAATGGATTGGATGGATTCATCCTGAG 516
           ||||||||||||||||||||||||||||||||||| ||||||||||||||| ||||||||
Db     502 CACTGGGTGAAGCAGAGGCCTGAACAGGGCCTGGAGTGGATTGGATGGATTGATCCTGAG 561

Qy     517 AATGGTAATACTGTATATGACCCGAAATTCCAGGGCAAGGCCAGTATAACAGCAGACACA 576
           ||||||  |||||  |||| |||||| |||||||||||||||| ||| || |||||||||
Db     562 AATGGTGGTACTGATTATGCCCCGAAGTTCCAGGGCAAGGCCACTATGACTGCAGACACA 621

Qy     577 TCCTCCAACGCGGCCTACCTTCAGCTCAGCAGCCTGACATCTGAGGACACTGCCGTCTAT 636
           ||||||||| | |||||||| |||||||||||||||||||||||||||||||||||||||
Db     622 TCCTCCAACACAGCCTACCTGCAGCTCAGCAGCCTGACATCTGAGGACACTGCCGTCTAT 681

Qy     637 TACTGTGCTTCTTATTACTACTATAGTGCTTAC 669
           ||||||  | | |||    ||||    |  |||
Db     682 TACTGTAATGCATATGGTAACTACGACGGCTAC 714


RESULT   9
MMZ95477
LOCUS       MMZ95477       822 bp    RNA             ROD       29-OCT-1997
DEFINITION  Mus musculus mRNA for IgG1/kappa antibody, scFv59-CK.
ACCESSION   Z95477 AF004404
NID         g2108310
KEYWORDS    IgG1.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 822)
  AUTHORS   Schouten,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-1997) Dept. of Nematology, Wageningen
            Agricultural University, P.O. Box 8123, Wageningen 6700 ES, The
            Netherlands
REFERENCE   2  (bases 1 to 822)
  AUTHORS   Schouten,A., Roosien,J., de Boer,J.M., Wilmink,A., Rosso,M.N.,
            Bosch,D., Stiekema,W.J., Gommers,F.J., Bakker,J. and Schots,A.
  TITLE     Improving scFv antibody expression levels in the plant cytosol
  JOURNAL   FEBS Lett. 415 (2), 235-241 (1997)
  MEDLINE   98010486
FEATURES             Location/Qualifiers
     source          1. .822
                     /organism="Mus musculus"
                     /note="recombinant gene"
                     /db_xref="taxon:10090"
     CDS             1. .822
                     /codon_start=1
                     /product="IgG1/kappa antibody"
                     /db_xref="PID:e1167935"
                     /db_xref="PID:g2108311"
                     /translation="MDVVMTQTPLSLPVSLGDQASISCRSSQTILHSNGNTYLEWFLQ
                     KPGQSPKLLIYKVSNRFSGVPDRFSGSGSGTDFTLKISRVEAEDLGVYYCFQGSHVPY
                     TFGGGTKLEIKGGSGGGGSGGGGSGGGGSEVQLQQSGAELVRSGASVKLSCTASDFNI
                     KDYYIHWVKQRPEQGLEWIGWIDPKNGDTDYAPKFQGKATMTADTSSNTAYLQLSSLT
                     SEDTAVYYCSAYGDFDAYRGQGTTLTVSSAKTTPGAAAEQKLISEEDLNDIKDEL"
     misc_feature    4. .339
                     /note="Light chain variable domain (Vl) coding sequence"
     misc_feature    340. .393
                     /note="Linker peptide coding sequence"
     misc_feature    394. .741
                     /note="Heavy chain variable domain (Vh) coding sequence"
     misc_feature    742. .756
                     /note="Heavy chain first constant domain (CH1) fragment
                     coding sequence"
     misc_feature    769. .801
                     /note="c-myc tag coding sequence"
     misc_feature    808. .819
                     /note="KDEL coding sequence"
BASE COUNT      214 a    203 c    221 g    184 t
ORIGIN


  Query Match             45.5%;  Score 327.6;  DB 13;  Length 822;
  Best Local Similarity   73.9%;  Pred. No. 1.5e-90;
  Matches  436;  Conservative    0;  Mismatches  139;  Indels   15;  Gaps    1;

Qy     95 ACTGGTATCAGCAGAAGTCAAGCACCTCCCCCAAACTCTGGGTTTATGACACATCCAAAC 154
          | ||||  | ||||||  || ||   || || || |||  | | ||  |    ||||| |
Db    119 AATGGTTCCTGCAGAAACCAGGCCAGTCTCCAAAGCTCCTGATCTACAAAGTTTCCAACC 178

Qy    155 TGGCTTCTGGAGTCCCAGGTCGCTTCAGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCA 214
              |||||| |||||||   | |||||||||||||| || || |    || | | ||||
Db    179 GATTTTCTGGGGTCCCAGACAGGTTCAGTGGCAGTGGATCAGGGACAGATTTCACACTCA 238

Qy    215 CGATCAGCAGCATGGAGGCTGAAGATGCTGCCACTTATTATTGTTATCAGGGGAGTGGGT 274
           |||||||||  |||||||||| |||   |    |||||| || | ||| ||
Db    239 AGATCAGCAGAGTGGAGGCTGAGGATCTGGGAGTTTATTACTGCTTTCAAGGTTCACATG 298

Qy    275 ACCCATTCACGTTCGGCTCGGGGACAAAGTTGGAAATAAAAGGTTCTACCTCCGGATCTG 334
            || | |||||||||   |||||||||| |||||||||||||     ||   ||   ||
Db    299 TTCCGTACACGTTCGGAGGGGGGACAAAGCTGGAAATAAAAGGGGGATCCGGTGGTGGTG 358

Qy    335 GTAAATCTTCTGAAGGTAAAGG---------------TGTGCAGCTGCAGCAGTCTGGGG 379
          ||     |  || |||    ||                || |||||||| ||||||||||
Db    359 GTTCCGGTGGTGGAGGCTCCGGAGGCGGAGGATCCGAGGTCCAGCTGCAACAGTCTGGGG 418

Qy    380 TTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGTCCTGCAAAGCTTCTGACTTCAACA 439
            ||||||||  |   |||||||| ||||||||||||||||| |||||||||||||||||
Db    419 CAGAGCTTGTGAGGTCAGGGGCCTCAGTCAAGTTGTCCTGCACAGCTTCTGACTTCAACA 478

Qy    440 TTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCCTGAACAGGGCCTGGAATGGATTG 499
          |||||||||| ||||| ||||||||||||||||||||||||||||||||||| |||||||
Db    479 TTAAAGACTACTATATACACTGGGTGAAGCAGAGGCCTGAACAGGGCCTGGAGTGGATTG 538

Qy    500 GATGGATTCATCCTGAGAATGGTAATACTGTATATGACCCGAAATTCCAGGGCAAGGCCA 559
          |||||||| ||||| |||||||| ||||||  |||| |||||| ||||||||||||||||
Db    539 GATGGATTGATCCTAAGAATGGTGATACTGATTATGCCCCGAAGTTCCAGGGCAAGGCCA 598

Qy    560 GTATAACAGCAGACACATCCTCCAACGCGGCCTACCTTCAGCTCAGCAGCCTGACATCTG 619
           ||| || |||||||||||||||||| | |||||||| ||||||||||||||||||||||
Db    599 CTATGACTGCAGACACATCCTCCAACACAGCCTACCTGCAGCTCAGCAGCCTGACATCTG 658

Qy    620 AGGACACTGCCGTCTATTACTGTGCTTCTTATTACTACTATAGTGCTTAC 669
          ||||||||||||||||||||||| | | |||    |||     || |||
Db    659 AGGACACTGCCGTCTATTACTGTAGTGCCTATGGTGACTTCGACGCCTAC 708
```

```
RESULT  10
E13598
LOCUS       E13598        735 bp    DNA             PAT       27-APR-1998
DEFINITION  DNA encoding a single chain Fv antibody of anti-T3 antibody TT1.
ACCESSION   E13598
NID         g3252400
KEYWORDS    JP 1997220092-A/1.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 735)
  AUTHORS   Ekida,T., Yasukawa,K., Imanaka,T. and Takagi,M. .
  TITLE     PRODUCTION OF SINGLE-STRAND FV ANTIBODY
  JOURNAL   Patent: JP 1997220092-A 1 26-AUG-1997;
            TOSOH CORP
COMMENT     OS   None
            OC   Artificial sequences.
            PN   JP 1997220092-A/1
            PD   26-AUG-1997
            PF   15-FEB-1996 JP 1996027622
            PI   EKIDA TEIJI, YASUKAWA KIYOSHI, IMANAKA TADAYUKI, PI   TAKAGI
            MASAHIRO
            PC   C12N15/09,C12N1/21,C12P21/08,(C12N1/21,C12R1:19),(C12P21/08,
            PC   C12R1:19);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .735
            FT                   /organism='Artificial sequences'.
FEATURES             Location/Qualifiers
     source          1. .735
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      165 a    192 c    197 g    181 t
ORIGIN


  Query Match             45.3%;  Score 326.4;  DB 6;  Length 735;
  Best Local Similarity   67.6%;  Pred. No. 3.6e-90;
  Matches  491;  Conservative    0;  Mismatches  226;  Indels    9;  Gaps    2;

Qy      1 ATGCAGCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATG 60
          || |||||||||||||||||| | | ||||  ||||||||| ||||| ||| |||| ||
Db     10 ATTCAGCTGACCCAGTCTCCAACCACCATGGTTGCATCTCCCGGGGAGAAGATCACTATC 69

Qy     61 ACCTGCAGTGCCAGCTCAAGTGTAAG------TAACATGCACTGGTATCAGCAGAAGTCA 114
          ||||||||||||||||||||| ||||      | || |||| ||||||||||||||| ||
Db     70 ACCTGCAGTGCCAGCTCAAGTATAAGTTCCAATTACTTGCATTGGTATCAGCAGAAGCCA 129

Qy    115 AGCACCTCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGT 174
           |   |||||| ||||||| | |||||   |||||||| ||||||||||| ||||||  |
Db    130 GGATTCTCCCCTAAACTCTTGATTTATAGGACATCCAATCTGGCTTCTGGTGTCCCAACT 189

Qy    175 CGCTTCAGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCT 234
          ||||||||||||||||||||||| | ||||||||||||||| ||  ||| ||||||||||
Db    190 CGCTTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATTGGCACCATGGAGGCT 249

Qy    235 GAAGATGCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCG 294
          ||||||| ||||||||| || ||  | ||||| ||| |    ||  ||||||||||  |
Db    250 GAAGATGTTGCCACTTACTACTGCCAGCAGGGTAGTAGTATACCGCTCACGTTCGGTGCT 309

Qy    295 GGGACAAAGTTGGAAATAAAAGGTTCTACCTCCGGATCTGGTAAATCTTCTGAAGGTAAA 354
          ||||| ||| | || ||    ||   |  ||| ||   ||||   ||   ||  ||   |
Db    310 GGGACCAAGCTCGAGATCGGTGGCGGTGGCTCGGGCGGTGGTGGGTCGGGTGGCGGCGGA 369

Qy    355 ---GGTGTGCAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAG 411
             |  ||  |||||||| |||||||||  |   | ||       || |  |   | ||
Db    370 TCCGAGGTCAAGCTGCAGGAGTCTGGGGGAGGCTTAGTGAAGCTTGGCGGGTCCCTGAAA 429

Qy    412 TTGTCCTGCAAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAG 471
```

```
            | |||||  |||| ||||  ||||   | |     ||||| ||    |||||    |||
Db     430 CTCTCCTGTGAAGCCTCTGGATTCACTTTCAGTAGTTATTACATGTCTTGGGTTCGCCAG 489

Qy     472 AGGCCTGAACAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTA 531
           |  || ||  || | ||||| | | | |   ||| ||  | |   |||||  ||    
Db     490 ACTCCAGAGAAGAGGCTGGAGTTGGTCGCAGCCATTAATAGTAATGGTGGTACCACCTAC 549

Qy     532 TATGACCCGAAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCC 591
           |||      |   |  |||||     ||  ||  |   |||||   ||   ||| |    
Db     550 TATTCAGACACTGTGAAGGGCCGATTCACCATCTCCAGAGACAATGCCAAGAACACCCTG 609

Qy     592 TACCTTCAGCTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGCTTCTTAT 651
           ||||| ||  | ||||| ||||  |||||||||||| ||| | |||||||||||       
Db     610 TACCTGCAAATGAGCAGTCTGAAGTCTGAGGACACAGCCTTGTATTACTGTGCAAGCCCG 669

Qy     652 TACTACTATAGTGCTTACTATGCTATGTACTACTGGGGTCAAGGAACCTCGGTCACCGTC 711
             || ||||  |     |||||  |||||||||||||| ||||| ||| ||||||||||
Db     670 GTCTCCTATTATTACCTCTATGTCATGTACTACTGGGGCCAAGGACCACGGTCACCGTC 729

Qy     712 TCCTCA 717
           ||||||
Db     730 TCCTCA 735


RESULT  11
I08678
LOCUS       I08678        741 bp                   PAT        14-NOV-1994
DEFINITION  Sequence 10 from Patent WO 8801649.
ACCESSION   I08678
NID         g588616
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 741)
  AUTHORS   Ladner,R.C. and Bird,R.E.
  TITLE     SINGLE POLYPEPTIDE CHAIN BINDING MOLECULES
  JOURNAL   Patent: WO 8801649-A 10 10-MAR-1988;
FEATURES             Location/Qualifiers
     source          1..741
                     /organism="unknown"
BASE COUNT      172 a    188 c    192 g    189 t
ORIGIN


  Query Match             44.6%;  Score 321.2;  DB 6;  Length 741;
  Best Local Similarity   69.0%;  Pred. No. 1.4e-88;
  Matches  504;  Conservative    0;  Mismatches  208;  Indels   18;  Gaps    4;

Qy       6 GCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATGACCTG 65
           ||| |||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db      12 GCTCACCCAGTCTCGAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATGACCTG 71

Qy      66 CAGTGCCAGCTCAAGTGTA------AGTAACATGCACTGGTATCAGCAGAAGTCAAGCAC 119
           ||| |||||||||||||||      ||| || ||||||||  |||||||||||| |  |
Db      72 CAGGGCCAGCTCAAGTGTAAGTTCCAGTTACTTGCACTGGTTCCAGCAGAAGTCAGGTGC 131

Qy     120 CTCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTT 179
           |||||||||||||||| |||||  |||||||||  | |||||||||||||| | ||||||
Db     132 CTCCCCCAAACTCTGGATTTATAGCACATCCAACTTCGCTTCTGGAGTCCCTGCTCGCTT 191

Qy     180 CAGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGA 239
           |||||||||||||||||| | |||||||||||||||| ||||||||  |||||||||||||
Db     192 CAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGTGTGGAGGCTGAAGA 251

Qy     240 TGCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGGAC 299
           ||||||||||||||| ||  |  |  |  |||| |||||| |||||||||||  | |||||
Db     252 TGCTGCCACTTATTACTGCCAGTACAGTGGTGGTTACCCACTCACGTTCGGTGCTGGGAC 311

Qy     300 AAAGTTGGAAATAAAAGGTTC---TACCTCCGGATCTGGTAAATCTTCTGAAGGTAAAGG 356
            ||| || ||  |   |||||   | | || | |             | |     |   |
Db     312 CAAGCTGAAAGAATCTGGTTCTGTTTCTTCTGAACAGCTGGCTCAGTTTCGTTCTCTGGA 371
```

```
Qy     357 TGTGCAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGTC 416
           ||||||||||  | |||||||||  ||  | ||       || |  |   | ||  | ||
Db     372 TGTGCAGCTGGTGGAGTCTGGGGGAGACTTAGTGAAGCCTGGAGGGTCCCTGAAACTCTC 431

Qy     417 CTGCAAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCC 476
           |||   ||| ||||  ||||   | |    ||||   ||    |||||    |  |  ||
Db     432 CTGTGCAGCCTCTGGATTCACTTTCATTAGCTATGGCATGTCTTGGGTTCGCCGAACTCC 491

Qy     477 TGAACAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATGA 536
            ||  || | ||||| ||| | | |   |||  |  ||  | |  | | ||    |||  
Db     492 AGACAAGAGGCTGGAGTGGGTCGCAACCATTAGTAGTGGTAGTACTTACACCTACTATCC 551

Qy     537 CCCGAAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTACCT 596
               |   |  ||||      ||  ||  |   |||||   ||   ||| |    |||||
Db     552 AGACAGTGTGAAGGGGCGATTCACCATCTCCAGAGACAATGCCAAGAACACCCTGTACCT 611

Qy     597 TCAGCTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGC--TTCTTATTAC 654
            ||  | ||| | ||||  |||||||||||| ||| | |||||||||||    |  |||||
Db     612 GCAAATGAGCGGTCTGAAGTCTGAGGACACAGCCATGTATTACTGTGCAAGACGGATTAC 671

Qy     655 TACTATAGT-------GCTTACTATGCTATGTACTACTGGGGTCAAGGAACCTCGGTCAC 707
           |||  ||||       | |||||||||||| ||||||||||||||||||||||||| |||||
Db     672 TACGGTAGTACTTACGGATTACTATGCTATGGACTACTGGGGTCAAGGAACCTCAGTCAC 731

Qy     708 CGTCTCCTCA 717
           |||||||| |
Db     732 CGTCTCCTAA 741


RESULT  12
I08679
LOCUS       I08679        753 bp                   PAT        14-NOV-1994
DEFINITION  Sequence 12 from Patent WO 8801649.
ACCESSION   I08679
NID         g588617
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 753)
  AUTHORS   Ladner,R.C. and Bird,R.E.
  TITLE     SINGLE POLYPEPTIDE CHAIN BINDING MOLECULES
  JOURNAL   Patent: WO 8801649-A 12 10-MAR-1988;
FEATURES             Location/Qualifiers
     source          1..753
                     /organism="unknown"
BASE COUNT      187 a    192 c    194 g    176 t      4 others
ORIGIN


  Query Match             43.3%;  Score 311.6;  DB 6;  Length 753;
  Best Local Similarity   67.9%;  Pred. No. 1.3e-85;
  Matches  504;  Conservative    0;  Mismatches  208;  Indels   30;  Gaps    4;

Qy       6 GCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATGACCTG 65
           ||| |||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db      12 GCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGCGGAAAAGGTCACCATGACCTG 71

Qy      66 CAGTGCCAGCTCAAGTGTA------AGTAACATGCACTGGTATCAGCAGAAGTCAAGCAC 119
           ||| |||||||||||||||      |||  ||   ||||||  |||||||||||| |  |
Db      72 CAGGGCCAGCTCAAGTGTAAGTTCCAGTTCCAGTTACTGGTTCCAGCAGAAGTCAGGTGC 131

Qy     120 CTCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTT 179
           |||||||||||||||| |||||  |||||||||  ||||||||||||||||| | ||||||
Db     132 CTCCCCCAAACTCTGGATTTATAGCACATCCAACTTGGCTTCTGGAGTCCCTGCTCGCTT 191

Qy     180 CAGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGA 239
           |||||||||||||||||| | |||||||||||||||| ||||||||  |||||||||||||
Db     192 CAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGTGTGGAGGCTGAAGA 251

Qy     240 TGCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGGAC 299
           ||||||||||||||| ||  | |||   |||| |||||| |||||||||||  | |||||
Db     252 TGCTGCCACTTATTACTGCCAGCAGTACAGTGGTTACCCACTCACGTTCGGTGCTGGGAC 311
```

```
Qy     300 AAAGTTGGAAATAAAAGGTTCTACCTCCGGATCTGGTAAATCTTCTGAAGGTAAAGGTGT 359
            ||| ||||   |.||||  ||   ||   | | | |||| ||   |       ||
Db     312 CAAGCTGGAGGCAGAAGGCACTCTGTCTCCAGCAGATAAAACTAACGTTAAAGCAGCATG 371

Qy     360 G---------------CAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTT 404
           |               ||||||  | |||||||||  ||  | ||       || |  |
Db     372 GGGCAAAGTTATGACTCAGCTGGTGGAGTCTGGGGGAGACTTAGTGAAGCCTGGAGGGTC 431

Qy     405 AGTCAAGTTGTCCTGCAAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGT 464
             | ||  | |||||   ||| ||||  ||||   | |    ||||   ||    |||||
Db     432 CCTGAAACTCTCCTGTGCAGCATCTGGATTCACTTTCATTAGCTATGGCATGTCTTGGGT 491

Qy     465 GAAGCAGAGGCCTGAACAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAA 524
               ||||  || ||  || | ||||| ||| | | |           ||  | |  | |
Db     492 TCGCCAGACTCCAGACAAGAGGCTGGAGTGGGTCGCAACCNNNAGNAGTGGTAGTACTTA 551

Qy     525 TACTGTATATGACCCGAAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAA 584
            ||    |||      |   |  ||||      ||  ||  |   |||||   ||   ||
Db     552 CACCTACTATCCAGACAGTGTGAAGGGGCGATTCACCATCTCCAGAGACAATGCCAAGAA 611

Qy     585 CGCGGCCTACCTTCAGCTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGC 644
           | |    ||||| ||  | ||| | ||||  ||||||||||| ||| | |||||||||||
Db     612 CACCCTGTACCTGCAAATGAGCGGTCTGAAGTCTGAGGACACAGCCATGTATTACTGTGC 671

Qy     645 --TTCTTATTACTACTATAGT-------GCTTACTATGCTATGTACTACTGGGGTCAAGG 695
               |  ||||||||  ||||       | ||||||||||||| ||||||||||||||||
Db     672 AAGACGGATTACTACGGTAGTACTTACGGATTACTATGCTATGGACTACTGGGGTCAAGG 731

Qy     696 AACCTCGGTCACCGTCTCCTCA 717
           |||||| ||||||||||||| |
Db     732 AACCTCAGTCACCGTCTCCTAA 753


RESULT  13
I31036
LOCUS       I31036         729 bp    DNA             PAT       20-DEC-1996
DEFINITION  Sequence 3 from patent US 5582826.
ACCESSION   I31036
NID         g1821827
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 729)
  AUTHORS   Shimamura,T., Hamuro,J., Nakazawa,H., Kanayama,Y., Sugamura,K. and
            Takeshita,T.
  TITLE     Monoclonal antibodies which bind the gamma chain of human
            interleukin-2 receptor
  JOURNAL   Patent: US 5582826-A 3 10-DEC-1996;
FEATURES             Location/Qualifiers
     source          1. .729
                     /organism="unknown"
BASE COUNT      199 a    169 c    182 g    179 t
ORIGIN


  Query Match             41.5%;  Score 299;  DB 6;  Length 729;
  Best Local Similarity   65.3%;  Pred. No. 1e-81;
  Matches  472;  Conservative    0;  Mismatches  245;  Indels    6;  Gaps    2;

Qy       1 ATGCAGCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATG 60
           || | |||||| |||||||||||   | | |||||||||   || ||||  ||||||||
Db       7 ATTCTGCTGACACAGTCTCCAGCCTCCCTATCTGCATCTGTGGGAGAAACTGTCACCATC 66

Qy      61 ACCTGCAGTGCCAGCTCAAGTGTA---AGTAACATGCACTGGTATCAGCAGAAGTCAAGC 117
           || ||  | || ||    | | |    | | |  |    ||||||||||||||     |
Db      67 ACATGTCGAGCAAGTGGGAATATTCACAATTATTTAGCATGGTATCAGCAGAAACAGGGA 126

Qy     118 ACCTCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGC 177
           |  || ||  | |||  ||| ||| |  ||   |   | ||   ||| || |||    |
Db     127 AAATCTCCTCAGCTCCTGGTCTATAATGCAAAAACCTTAGCAGATGGTGTGCCATCAAGG 186
```

```
Qy     178 TTCAGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAA 237
           |||||||||||||| || ||||     || ||||||| ||||| |||| || || |||||
Db     187 TTCAGTGGCAGTGGATCAGGAACACAATATTCTCTCAAGATCAACAGCCTGCAGCCTGAA 246

Qy     238 GATGCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGG 297
           |||  ||  | |||||| ||| | ||     |  |    || |  ||||||||    |||
Db     247 GATTTTGGGAGTTATTACTGTCAACATTTTTGGAGTACTCCGTGGACGTTCGGTGGAGGG 306

Qy     298 ACAAAGTTGGAAATAAAAGGTTCTA---CCTCCGGATCTGGTAAATCTTCTGAAGGTAAA 354
           || ||| ||||  | ||||     |   |||| ||||||||       ||  || | |
Db     307 ACCAAGCTGGAGCTCAAAGTCGAGAAATCCTCAGGATCTGGCTCCGAATCCAAAAGCACG 366

Qy     355 GGTGTGCAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTG 414
              ||  | ||  || |||||||   |||||| ||       ||||| | ||| ||| |
Db     367 CAGGTCAAACTCGAGGAGTCTGGACCTGAGCTGGTGAAGCCTGGGGCTTCAGTGAAGATA 426

Qy     415 TCCTGCAAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGG 474
           |||||||| |||||||  | |    | |  | ||| || || ||||||||||||||.||
Db     427 TCCTGCAAGGCTTCTGGTTACTCATTCACTGGCTACTACATGCACTGGGTGAAGCAAAGC 486

Qy     475 CCTGAACAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATAT 534
           | || | || |||| || ||||||||| | ||| ||||| | ||||||  ||||   ||
Db     487 CATGTAAAGAGCCTTGAGTGGATTGGACGTATTAATCCTTACAATGGTGCTACTAGCTAC 546

Qy     535 GACCCGAAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTAC 594
            ||| ||| ||| ||| |||||||||  | || | ||| |  ||||||| | | ||||||
Db     547 AACCAGAATTTCAAGGACAAGGCCAGCTTGACTGTAGATAAGTCCTCCAGCACAGCCTAC 606

Qy     595 CTTCAGCTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGCTTCTTATTAC 654
            |  |||||  ||||||||||||||||||| |||| ||||||||||||||     |  |
Db     607 ATGGAGCTCCACAGCCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGAGAGAAT 666

Qy     655 TACTATAGTGCTTACTATGCTATGTACTACTGGGGTCAAGGAACCTCGGTCACCGTCTCC 714
           |||||  ||  |  ||| |   |    |||||||| ||||| ||   |||||| |||||
Db     667 TACTACGGTAGTAGCTACGGGTTTGCTTACTGGGGCCAAGGGACTCTGGTCACTGTCTCT 726

Qy     715 TCA 717
            ||
Db     727 GCA 729


RESULT  14
E10362
ID   E10362     standard; RNA; ROD; 729 BP.
XX
AC   E10362;
XX
NI   d1108699
XX
DT   08-OCT-1997 (Rel. 52, Created)
DT   08-OCT-1997 (Rel. 52, Last updated, Version 1).
XX
DE   cDNA encoding an monoclonal antibody against human Interleukin-2
DE   receptor gamma chain.
XX
KW   JP 1995313188-A/2.
XX
OS   Mus sp.
OC   Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Rodentia;
OC   Sciurognathi; Muridae; Murinae; Mus.
XX
RN   [1]
RP   1-729
RA   Shimamura T., Hamuro J., Nakazawa H., Kanayama Y., Sugamura K.,
RA   Takeshita T.;
RT   "IMMUNOSUPPRESSIVE AGENT";
RL   Patent number JP 1995313188-A/2, 05-DEC-1995.
RL   AJINOMOTO CO INC, SUGAMURA KAZUO,.
XX
CC   OS   Mus sp. (mouse)
CC   PN   JP 1995313188-A/2
CC   PD   05-DEC-1995
CC   PF   21-APR-1994 JP 1994082836
CC   PR    21-APR-1993 JP 93P    94491, 07-MAR-1994 JP 94P    36065
CC   PI    SHIMAMURA TOSHIAKI, HAMURO JUNJI, NAKAZAWA HARUMI,
CC   PI    KANAYAMA YUKA,
CC   PI    SUGAMURA KAZUO, TAKESHITA TOSHIICHI
CC   PC    C12P21/08,A61K39/395,A61K39/395,C12N1/21,C12N5/20,
CC   PC    C12N15/13//C12N15/06,
CC   PC    (C12P21/08,C12R1:19),(C12P21/08,C12R1:91),(C12N1/21,C12R1:19);
CC   FH    Key             Location/Qualifiers
CC   FH
CC   FT    source          1. .729
CC   FT                    /organism="Mus sp."
CC   FT                    /cell_type="hybridoma"
CC   FT                    /cell_line="GP-4"
CC   FT    mat_peptide     1. .729
CC   FT                    /product="anti-IL-2 receptor gamma chain"
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .729
FT                   /organism="Mus sp."
XX
SQ   Sequence 729 BP; 199 A; 169 C; 182 G; 179 T; 0 other;


  Query Match             41.5%;  Score 299;  DB 27;  Length 729;
  Best Local Similarity   65.3%;  Pred. No. 1e-81;
  Matches  472;  Conservative    0;  Mismatches  245;  Indels    6;  Gaps    2;

Qy      1 ATGCAGCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATG 60
          || | |||||| |||||||||||   | | |||||||||   || ||||  ||||||||
Db      7 ATTCTGCTGACACAGTCTCCAGCCTCCCTATCTGCATCTGTGGGAGAAACTGTCACCATC 66

Qy     61 ACCTGCAGTGCCAGCTCAAGTGTA---AGTAACATGCACTGGTATCAGCAGAAGTCAAGC 117
          || ||  | || ||    | | |    | | |  |    ||||||||||||||     |
Db     67 ACATGTCGAGCAAGTGGGAATATTCACAATTATTTAGCATGGTATCAGCAGAAACAGGGA 126

Qy    118 ACCTCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGC 177
          |  || ||  | |||  ||| ||| |  ||   |   | ||   ||| || |||    |
Db    127 AAATCTCCTCAGCTCCTGGTCTATAATGCAAAAACCTTAGCAGATGGTGTGCCATCAAGG 186

Qy    178 TTCAGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAA 237
          |||||||||||||| || ||||     || ||||||| ||||| |||| || || |||||
Db    187 TTCAGTGGCAGTGGATCAGGAACACAATATTCTCTCAAGATCAACAGCCTGCAGCCTGAA 246

Qy    238 GATGCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGG 297
          |||  ||  | |||||| ||| | ||     |  |    || |  ||||||||    |||
Db    247 GATTTTGGGAGTTATTACTGTCAACATTTTTGGAGTACTCCGTGGACGTTCGGTGGAGGG 306

Qy    298 ACAAAGTTGGAAATAAAAGGTTCTA---CCTCCGGATCTGGTAAATCTTCTGAAGGTAAA 354
          || ||| ||||  | ||||     |   |||| ||||||||       ||  || | |
Db    307 ACCAAGCTGGAGCTCAAAGTCGAGAAATCCTCAGGATCTGGCTCCGAATCCAAAAGCACG 366

Qy    355 GGTGTGCAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTG 414
             ||  | ||  || |||||||   |||||| ||       ||||| | ||| ||| |
Db    367 CAGGTCAAACTCGAGGAGTCTGGACCTGAGCTGGTGAAGCCTGGGGCTTCAGTGAAGATA 426

Qy    415 TCCTGCAAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGG 474
          |||||||| |||||||  | |    | |  | ||| || || |||||||||||||| ||
Db    427 TCCTGCAAGGCTTCTGGTTACTCATTCACTGGCTACTACATGCACTGGGTGAAGCAAAGC 486

Qy    475 CCTGAACAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATAT 534
          | || | || |||| || ||||||||| | ||| ||||| | ||||||  ||||   ||
Db    487 CATGTAAAGAGCCTTGAGTGGATTGGACGTATTAATCCTTACAATGGTGCTACTAGCTAC 546

Qy    535 GACCCGAAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTAC 594
           ||| ||| ||| ||| |||||||||  | || | ||| |  ||||||| | | ||||||
Db    547 AACCAGAATTTCAAGGACAAGGCCAGCTTGACTGTAGATAAGTCCTCCAGCACAGCCTAC 606

Qy    595 CTTCAGCTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGCTTCTTATTAC 654
           |  |||||  ||||||||||||||||||| |||| ||||||||||||||     |  |
Db    607 ATGGAGCTCCACAGCCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGAGAGAAT 666

Qy    655 TACTATAGTGCTTACTATGCTATGTACTACTGGGGTCAAGGAACCTCGGTCACCGTCTCC 714
```

```
          |||||  ||  |  ||| |   |    |||||||| ||||| ||   |||||| |||||
Db    667 TACTACGGTAGTAGCTACGGGTTTGCTTACTGGGGCCAAGGGACTCTGGTCACTGTCTCT 726

Qy    715 TCA 717
           ||
Db    727 GCA 729


RESULT  15
SYNDA44X
LOCUS       SYNDA44X      744 bp    mRNA            SYN       01-MAR-1996
DEFINITION  Synthetic single-chain Fv fusion protein (DA4.4/212) mRNA,
            constructed from variable light and heavy chain regions of mouse
            monoclonal antibody DA4.4, 3' end of cds.
ACCESSION   L17037
NID         g310739
KEYWORDS    IgM-binding protein; fusion protein; fusion protein DA4.4/212;
            monoclonal antibody.
SOURCE      Artificial gene cDNA to mRNA.
  ORGANISM  artificial sequence
            artificial sequence.
REFERENCE   1  (bases 1 to 744)
  AUTHORS   Lee,T.K., Rollence,M.L., Hallberg,P.L., Oelkuct,M.S., Dodd,S.W.,
            Nagle,J.W. and Filpula,D.R.
  TITLE     Production of engineered IgM-binding single-chain antibodies in
            Escherichia coli
  JOURNAL   J. Ind. Microbiol. 14 (5), 371-376 (1995)
  MEDLINE   95336669
FEATURES             Location/Qualifiers
     source          1. .744
                     /organism="artificial sequence"
                     /db_xref="taxon:29278"
     CDS             <1. .735
                     /standard_name="DA4.4/212 sFv protein"
                     /codon_start=1
                     /transl_table=11
                     /function="binds human IgM"
                     /product="fusion protein"
                     /db_xref="PID:g310740"
                     /translation="DVVMTQSPSSLAMSVGQKVTMSCKSSQSLLNSSNQKNYLAWYQQ
                     KPGQSPELLVYFASTRESGVPDRFIGSGSGTDFTLTISSVQAEDLADYFCQQHYSTPF
                     TFGSGTKLEIKGSTSGSGKSSEGKGQVQLQQPGAEFVKPGAPVKLSCKASGYPFTTYW
                     VNWMKQRPGRGLEWIGRIDPYDSETLYNQKFKDKATLTVDKSSSTAYIQLSSLTSEDS
                     AVYYCARETYDYPFAYWGQGTLVTVS"
     misc_feature    1. .339
                     /note="variable light chain derived from mouse monoclonal
                     antibody DA4.4 (ATCC HB57)"
     misc_feature    340. .381
                     /note="linker segment"
     misc_feature    382. .744
                     /note="variable heavy chain derived from mouse monoclonal
                     antibody DA4.4 (ATCC HB57)"
BASE COUNT      192 a    187 c    186 g    179 t
ORIGIN


  Query Match             41.3%;  Score 297.4;  DB 15;  Length 744;
  Best Local Similarity   66.3%;  Pred. No. 3.2e-81;
  Matches  490;  Conservative    0;  Mismatches  216;  Indels   33;  Gaps    3;

Qy      6 GCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATGACCTG 65
          | |||| ||||||||| |   | || ||   ||   |||  | |||||||| ||||  ||
Db      9 GATGACACAGTCTCCATCCTCCCTGGCTATGTCAGTAGGACAGAAGGTCACTATGAGTTG 68

Qy     66 CAGTGCCAGCTCAAGTGTAAGTAACA---------------------TGCACTGGTATCA 104
          ||   ||||    ||| |    || |                     ||  |||||| ||
Db     69 CAAGTCCAGTCAGAGTCTTTTAAATAGTAGCAATCAAAAGAACTATTTGGCCTGGTACCA 128

Qy    105 GCAGAAGTCAAGCACCTCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGG 164
          ||||||  || |    || ||  ||||   ||| ||     ||||||    ||  |||||
Db    129 GCAGAAACCAGGACAGTCTCCTGAACTTCTGGTATACTTTGCATCCACTAGGGAATCTGG 188

Qy    165 AGTCCCAGGTCGCTTCAGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAG 224
```

```
            ||||| | ||||||||  |||||||| ||||| |    || | |||| || ||||||||
Db     189 GGTCCCTGATCGCTTCATAGGCAGTGGATCTGGGACAGATTTCACTCTTACCATCAGCAG 248

Qy     225 CATGGAGGCTGAAGATGCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCAC 284
             || ||||||||||    ||   ||| |  ||| | ||      | |    ||||||||
Db     249 TGTGCAGGCTGAAGACCTGGCAGATTACTTCTGTCAGCAACATTATAGCACTCCATTCAC 308

Qy     285 GTTCGGCTCGGGGACAAAGTTGGAAATAAAAGGTTCTACCTCCGGATCTGGTAAATCTTC 344
           ||||||||||||||||||| | ||||||||||||| ||||| |||||    || ||||| ||
Db     309 GTTCGGCTCGGGGACAAAGCTTGAAATAAAAGGCTCTACTTCCGGTAGCGGCAAATCCTC 368

Qy     345 TGAAGGTAAAGGT---GTGCAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGC 401
           |||||| ||||||   || ||||||||||||| |||||| |||| ||||       |||||
Db     369 TGAAGGCAAAGGTCAGGTTCAGCTGCAGCAGCCTGGGGCTGAGTTTGTGAAGCCTGGGGC 428

Qy     402 CTTAGTCAAGTTGTCCTGCAAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTG 461
              ||| ||  ||||||||| | ||||||| || |  | | |    ||| |   |  ||||
Db     429 TCCAGTGAAACTGTCCTGCAAGGCTTCTGGCTACCCCTTCACCACCTACTGGGTGAACTG 488

Qy     462 GGTGAAGCAGAGGCCTGAACAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGG 521
           | |||||||| |||||| ||  ||||| || ||||||||| ||||| |||||| |  || |
Db     489 GATGAAGCAGCGGCCTGGACGAGGCCTCGAGTGGATTGGAAGGATTGATCCTTACGATAG 548

Qy     522 TAATACTGTATATGACCCGAAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTC 581
           | | ||| | ||  | |  || ||| |||||||||    | || |  |||| ||||||
Db     549 TGAAACTCTCTACAATCAAAAGTTCAAGGACAAGGCCACACTGACTGTGGACAAATCCTC 608

Qy     582 CAACGCGGCCTACCTTCAGCTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTG 641
           || | | |||||| | || |||||||||||||||||||||||| |||| |||||||||||
Db     609 CAGCACAGCCTACATCCAACTCAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTG 668

Qy     642 TGCTTCTTATTACTACTATAGTGCTTACTATGCTATGTACTACTGGGGTCAAGGAACCTC 701
           |||              | | |  | | ||  |  |    |||||||| |||| ||
Db     669 TGC---------AAGGGAAACTTATGATTACCCCTTTGCTTACTGGGGCCAAGGGACTCT 719

Qy     702 GGTCACCGTCTCCTCATAA 720
           |||||| || || | ||||
Db     720 GGTCACTGTGTCTTAATAA 738


Search completed: May 13, 1999, 09:54:20
Job time: 1177 sec
```

```
                     GenCore version 4.5
             Copyright (c) 1993 - 1998  Compugen Ltd.


OM nucleic - nucleic search, using sw model

Run on:         May 13, 1999, 09:40:43 ; Search time 341.74 Seconds
                                         (without alignments)
                                         3213.863 Million cell updates/sec

Title:          US-08-704-178-2
Perfect score:  720
Sequence:       1 ATGCAGCTGACCCAGTCTCC..........CGGTCACCGTCTCCTCATAA 720

Scoring table:  IDENTITY_NUC

Searched:       2002476 seqs, 762712212 residues

Database :      EST:*
                1:  em_est1:*
                2:  em_est2:*
                3:  em_est3:*
                4:  em_est4:*
                5:  em_est5:*
                6:  em_est6:*
                7:  em_est7:*
                8:  em_est8:*
                9:  em_est9:*
                10:  gb_est1:*
                11:  gb_est2:*
                12:  gb_est3:*
                13:  gb_est4:*
                14:  gb_est5:*
                15:  gb_est6:*
                16:  gb_est7:*
                17:  gb_est8:*
                18:  gb_est9:*
                19:  gb_est10:*
                20:  gb_est11:*
                21:  gb_est12:*
                22:  gb_est13:*
                23:  gb_est22:*
                24:  gb_est14:*
                25:  gb_est15:*
                26:  gb_est16:*
                27:  gb_est17:*
                28:  gb_est18:*
                29:  gb_est19:*
                30:  gb_est20:*
                31:  gb_est21:*
                32:  em_est10:*
                33:  em_est11:*
                34:  em_est12:*
                35:  em_est13:*
                36:  em_est14:*
                37:  em_est15:*
                38:  em_est16:*
                39:  em_est17:*
                40:  em_est18:*
                41:  em_est19:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                 SUMMARIES
                     %
Result               Query
  No.    Score   Match Length DB   ID                     Description
-----------------------------------------------------------------------------
    1    251.2    34.9    396  25  AA691311               AA691311 vs14f01.r
    2    188.4    26.2    344  22  AA592800               AA592800 vo25g11.r
    3    149.8    20.8    310  30  AI180569               AI180569 uc70f09.r
c   4    146.2    20.3    461  31  AI201426               AI201426 qs73c06.x
    5    136.6    19.0    387  22  AA569186               AA569186 nm30d10.s
    6    133.8    18.6    294  28  AI007196               AI007196 ua73g04.r
    7    122.8    17.1    400  20  AA456778               AA456778 zw27f11.r
    8    122.2    17.0    393  11  T27593                 T27593 EST100653 H
    9      119    16.5    328  17  AA295311               AA295311 EST100471
   10      119    16.5    345  25  AA710291               AA710291 vt53a04.r
   11    118.8    16.5    403  18  AA345486               AA345486 EST51505
   12    117.6    16.3    370  17  AA295093               AA295093 EST100400
   13    117.6    16.3    333  17  AA300732               AA300732 EST13847
   14    116.6    16.2    379  21  AA291381               AA291381 zt44g02.r
   15    115.4    16.0    418  21  AA515239               AA515239 ng69c07.s
   16    113.8    15.8    363  18  AA367405               AA367405 EST78511
c  17    113.4    15.8    471  31  AI233978               AI233978 EST230666
   18      113    15.7    463  11  R69532                 R69532 yj82d09.r1
   19    112.8    15.7    349  18  AA379044               AA379044 EST91999
   20    112.6    15.6    335  17  AA318377               AA318377 EST20620
   21    112.4    15.6    398  24  AA423447               AA423447 ve80a03.r
   22    111.4    15.5    413  17  AA301347               AA301347 EST14279
   23    110.6    15.4    345  18  AA335086               AA335086 EST39457
   24    109.2    15.2    253  17  AA295941               AA295941 EST101165
   25    106.4    14.8    420  30  AI120005               AI120005 uc25e04.r
   26      106    14.7    381  11  T29114                 T29114 EST69430 Ho
   27      106    14.7    382  17  AA295786               AA295786 EST100987
   28    105.6    14.7    352  17  AA301261               AA301261 EST14181
   29    104.6    14.5    516  11  R67559                 R67559 yi42h11.r1
   30    104.6    14.5    346  17  AA300582               AA300582 EST13427
   31    104.2    14.5    823  16  AA170256               AA170256 ms87g10.r
   32    103.4    14.4    395  11  T29112                 T29112 EST69384 Ho
   33    102.6    14.2    470  12  H62115                 H62115 yu40h01.r1
   34      102    14.2    383  11  T29916                 T29916 EST99871 Ho
   35      101    14.0    405  18  AA360223               AA360223 EST69341
   36     97.8    13.6    303  17  AA300891               AA300891 EST14031
   37     96.8    13.4    339  25  AA710249               AA710249 vt49e01.r
   38     96.2    13.4    357  18  AA361497               AA361497 EST71040
   39       96    13.3    303  17  AA300788               AA300788 EST13648
   40     95.8    13.3    334  17  AA327254               AA327254 EST30547
   41     95.4    13.2    210  11  R69482                 R69482 yj83c03.r1
   42       95    13.2    372  25  AA581192               AA581192 nd38b11.r
   43     94.8    13.2    432  31  AI268604               AI268604 qo47a10.x
   44       94    13.1    431  25  AA710970               AA710970 vt93h09.r
   45     93.2    12.9    332  11  T29342                 T29342 EST77181 Ho


                                ALIGNMENTS


RESULT   1
AA691311
LOCUS       AA691311      396 bp    mRNA            EST       16-DEC-1997
DEFINITION  vs14f01.r1 Barstead mouse irradiated colon MPLRB7 Mus musculus cDNA
            clone 1138201 5' similar to gb:X67211 M.musculus rearranged
            immunoglobulin kappa light chain (MOUSE);, mRNA sequence.
ACCESSION   AA691311
NID         g2692247
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 396)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:619473
            Seq primer: -28m13 rev2 ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .396
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; Tissue obtained
                     from 8 week old mouse.  Colon was harvested 72 hours after
                     irradiation with 1400 Gys.  1st strand cDNA was primed
                     with  a Not I - oligo(dT) primer
                     [5'TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTT
                     T 3'];  double-stranded cDNA was ligated to Eco RI
                     adaptors [AATTCGGATCCTTG],  digested with Not I and cloned
                     into the Not I and Eco RI sites of  the modified pT7T3
                     vector. Library constructed by Bob Barstead."
                     /db_xref="taxon:10090"
                     /clone="1138201"
                     /clone_lib="Barstead mouse irradiated colon MPLRB7"
                     /dev_stage="8 weeks"
                     /lab_host="DH10B"
BASE COUNT      104 a    107 c     91 g     94 t
ORIGIN


  Query Match             34.9%;  Score 251.2;  DB 25;  Length 396;
  Best Local Similarity   90.5%;  Pred. No. 2.2e-68;
  Matches  268;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy        7 CTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATGACCTGC 66
            || |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db      101 CTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAGAAGGTCACCATGACCTGC 160

Qy       67 AGTGCCAGCTCAAGTGTAAGTAACATGCACTGGTATCAGCAGAAGTCAAGCACCTCCCCC 126
            ||||||||||||||||||||| ||||||||||||| |||||||||||| |||||||||||
Db      161 AGTGCCAGCTCAAGTGTAAGTTACATGCACTGGTACCAGCAGAAGTCAGGCACCTCCCCC 220

Qy      127 AAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTTCAGTGGC 186
            |||   ||| |||||||||||||||||||||||||||||||||| | |||||||||||||
Db      221 AAAAGATGGATTTATGACACATCCAAACTGGCTTCTGGAGTCCCTGCTCGCTTCAGTGGC 280

Qy      187 AGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGATGCTGCC 246
            ||||||||||| | ||||||||||||||| ||||||||||||||||||||||||||||||
Db      281 AGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGCATGGAGGCTGAAGATGCTGCC 340

Qy      247 ACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGGACAAA 302
            |||||||| ||  | ||| ||||| |  ||||  ||||||||||  | ||||| ||
Db      341 ACTTATTACTGCCAGCAGTGGAGTAGTAACCCGCTCACGTTCGGTGCTGGGACCAA 396


RESULT   2
AA592800
LOCUS       AA592800       344 bp    mRNA             EST       17-SEP-1997
DEFINITION  vo25g11.r1 Barstead mouse irradiated colon MPLRB7 Mus musculus cDNA
            clone 1050980 5' similar to gb:X67211 M.musculus rearranged
            immunoglobulin kappa light chain (MOUSE);, mRNA sequence.
ACCESSION   AA592800
NID         g2406463
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 344)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
```

```
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:582556
            Trace considered overall poor quality
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .344
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; Tissue obtained
                     from 8 week old mouse.  Colon was harvested 72 hours after
                     irradiation with 1400 Gys.  1st strand cDNA was primed
                     with  a Not I - oligo(dT) primer
                     [5'TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTT
                     T 3'];  double-stranded cDNA was ligated to Eco RI
                     adaptors [AATTCGGATCCTTG],  digested with Not I and cloned
                     into the Not I and Eco RI sites of  the modified pT7T3
                     vector. Library constructed by Bob Barstead."
                     /db_xref="taxon:10090"
                     /clone="1050980"
                     /clone_lib="Barstead mouse irradiated colon MPLRB7"
                     /dev_stage="8 weeks"
                     /lab_host="DH10B"
     mRNA            <1. .>344
BASE COUNT       91 a     91 c     70 g     92 t
ORIGIN


  Query Match             26.2%;  Score 188.4;  DB 22;  Length 344;
  Best Local Similarity   84.0%;  Pred. No. 9.9e-49;
  Matches  225;  Conservative    0;  Mismatches   41;  Indels    2;  Gaps    1;

Qy        7 CTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAA--AAGGTCACCATGACCT 64
            || || || |||||||||||||||||||| ||||||| || |  || ||||||| |||||
Db       75 CTCACTCAATCTCCAGCAATCATGTCTGCTTCTCCAGAGGGAGCAATGTCACCACGACCT 134

Qy       65 GCAGTGCCAGCTCAAGTGTAAGTAACATGCACTGGTATCAGCAGAAGTCAAGCACCTCCC 124
            |||||||||| || | ||||||| ||||||| ||||| |||||||||||| | |||||||
Db      135 GCAGTGCCAGTTCTACTGTAAGTTACATGCAATGGTAACAGCAGAAGTCAGGTACCTCCC 194

Qy      125 CCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTTCAGTG 184
            |||||   ||| ||||||| ||||| ||||||||||||||| |||| | |||||||||||
Db      195 CCAAAACATGGATTTATGATACATCTAAACTGGCTTCTGGAATCCCTGCTCGCTTCAGTG 254

Qy      185 GCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGATGCTG 244
            ||||| | ||||| | ||| ||||||||||| |||||||| || ||||||||| ||||||
Db      255 GCAGTAGCTCTGGGACCTCCTACTCTCTCACAATCAGCAGAATTGAGGCTGAACATGCTG 314

Qy      245 CCACTTATTATTGTTATCAGGGGAGTGG 272
            |||||||||| ||  | ||  ||||| |
Db      315 CCACTTATTACTGCCAGCATTGGAGTAG 342


RESULT   3
AI180569
LOCUS       AI180569       310 bp    mRNA             EST       08-OCT-1998
DEFINITION  uc70f09.r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone
            1431017 5' similar to gb:M18512 IG HEAVY CHAIN PRECURSOR V-I REGION
            (HUMAN); gb:M12376 Mouse immunoglobulin H-chain V-region pseudogene
            mRNA, complete (MOUSE);, mRNA sequence.
ACCESSION   AI180569
NID         g3731207
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 310)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:915085
            Seq primer: -28m13 rev2 ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .310
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
                     /db_xref="taxon:10090"
                     /clone="1431017"
                     /clone_lib="Soares mouse mammary gland NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
BASE COUNT       80 a     77 c     85 g     67 t      1 others
ORIGIN

  Query Match             20.8%;  Score 149.8;  DB 30;  Length 310;
  Best Local Similarity   72.5%;  Pred. No. 1.2e-36;
  Matches  208;  Conservative    0;  Mismatches   77;  Indels    2;  Gaps    1;

Qy     358 GTGCAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGTCC 417
           || || ||||||||| |||||||  |   |     ||   |||  | ||| ||| |||||
Db       4 GTCCAACTGCAGCAGCCTGGGGT--AAGCTTGTGAAGCCTGGGGTTCAGTGAAGCTGTCC 61

Qy     418 TGCAAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCCT 477
           ||||| ||||||| || || | | |    ||| |  || || ||||| || |||||||||
Db      62 TGCAAGGCTTCTGGCTACACCTTCACCAGCTACTGGATGCAGTGGGTAAAACAGAGGCCT 121

Qy     478 GAACAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATGAC 537
           | ||||||| | || ||||| |||  |||| |||||    || |  |||||   ||  |
Db     122 GGACAGGGCATTGAGTGGATCGGAGAGATTGATCCTTCTGATAGCTATACTAACTACAAT 181

Qy     538 CCGAAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTACCTT 597
           |  || ||| ||||||||||     | || | |||||||||||||| | | |||||| |
Db     182 CAAAAGTTCAAGGGCAAGGCAGCATTGACTGTAGACACATCCTCCAGCACAGCCTACATG 241
```

```
Qy     598 CAGCTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGC 644
           ||||||||||||||||||||||||||| |||| |||||| |||||||
Db     242 CAGCTCAGCAGCCTGACATCTGAGGACTCTGCGGTCTATGACTGTGC 288


RESULT     4
AI201426/c
LOCUS       AI201426      461 bp    mRNA            EST       30-OCT-1998
DEFINITION  qs73c06.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:1943722 3'
            similar to gb:M18512 IG HEAVY CHAIN PRECURSOR V-I REGION
            (HUMAN);contains Alu repetitive element;, mRNA sequence.
ACCESSION   AI201426
NID         g3754032
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            unknown library type
            Insert Length: 608   Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Homo sapiens"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr22 was prepared, and ss
                     circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     985608-986759, 1101192-1101959, and 1217928-1220615).
                     Subtraction by Bento Soares and M. Fatima Bonaldo.  "
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1943722"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
BASE COUNT       93 a    132 c    106 g    130 t
ORIGIN

  Query Match             20.3%;  Score 146.2;  DB 31;  Length 461;
  Best Local Similarity   69.3%;  Pred. No. 1.8e-35;
  Matches  199;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy     358 GTGCAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGTCC 417
           |||||||||  ||||||||||| |||| |          ||||||| ||| ||| | |||
Db     435 GTGCAGCTGGTGCAGTCTGGGGCTGAGGTGAAGAAGCCTGGGGCCTCAGTGAAGGTCTCC 376

Qy     418 TGCAAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCCT 477
           ||||| |||||||  | || | | |  ||||| ||||| |||||||||   |||   |||
Db     375 TGCAAGGCTTCTGGATACATCTTCACCGACTACTATATGCACTGGGTGCGACAGGCCCCT 316

Qy     478 GAACAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATGAC 537
           | ||| |  || |  ||||| ||| ||||  | ||| | | ||||   ||    ||||
Db     315 GGACAAGAGCTTGGGTGGATGGGACGGATCAACCCTAACAGTGGTGGCACAAACTATGCA 256

Qy     538 CCGAAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTACCTT 597
           | ||| || |||||||  | ||  || ||    ||||| |||  || | | ||||||
Db     255 CAGAAGTTTCAGGGCAGAGTCACCATGACCAGGGACACGTCCATCAGCACAGCCTACACG 196
```

```
Qy    598 CAGCTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGC 644
           |||| |||||||||| |||||||||||| |||   |||||||||||
Db    195 GAGCTGAGCAGCCTGAGATCTGAGGACACGGCCACGTATTACTGTGC 149


RESULT   5
AA569186
LOCUS       AA569186       387 bp    mRNA            EST       09-SEP-1997
DEFINITION  nm30d10.s1 NCI_CGAP_Lip2 Homo sapiens cDNA clone IMAGE:1061683
            similar to gb:L02325 IG HEAVY CHAIN PRECURSOR V-I REGION (HUMAN);,
            mRNA sequence.
ACCESSION   AA569186
NID         g2342240
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 387)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 1722   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 284.
FEATURES             Location/Qualifiers
     source          1. .387
                     /organism="Homo sapiens"
                     /note="Vector: pAMP10; mRNA made from liposarcoma, cDNA
                     made by oligo-dT priming. Non- directionally cloned.
                     Size-selected on agarose gel, average insert  size 600 bp.
                     Reference:  Krizman et al. (1996) Cancer Research
                     56:5380-5383."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1061683"
                     /clone_lib="NCI_CGAP_Lip2"
                     /tissue_type="liposarcoma"
                     /lab_host="DH10B"
     mRNA            <1. .>387
BASE COUNT       89 a     95 c    118 g     85 t
ORIGIN


  Query Match             19.0%;  Score 136.6;  DB 22;  Length 387;
  Best Local Similarity   67.2%;  Pred. No. 1.7e-32;
  Matches  193;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy    358 GTGCAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGTCC 417
          || || |||  ||||||||||| |||| |          ||||||| ||| ||| | |||
Db     98 GTCCAACTGGTGCAGTCTGGGGCTGAGGTGAAGAAGCCTGGGGCCTCAGTGAAGCTTTCC 157

Qy    418 TGCAAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCCT 477
          || || || ||||  |||| | | |   |||   |||| |||||||||   |||   |||
Db    158 TGTAAGGCATCTGGATTCACCCTCACCAACTGCCATATGCACTGGGTGCGACAGGCCCCT 217

Qy    478 GAACAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATGAC 537
          |  || || || || ||| | |||  |||  || ||    ||||| |||      | |
Db    218 GGGCAAGGGCTTGAGTGGGTGGGAATGATCAATTCTAGTGATGGTTATATAAGTAACGCA 277
```

```
Qy    538 CCGAAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTACCTT 597
          | ||| ||||||||||  | ||  || ||    ||||| ||| | | | | | |||  |
Db    278 CAGAAGTTCCAGGGCAGAGTCACCATGACCAGGGACACGTCCACGAGCACAGTCTATATG 337

Qy    598 CAGCTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGC 644
           |||| |||||||||| |||||||||||| |||||||||||||||||
Db    338 GAGCTTAGCAGCCTGAGATCTGAGGACACGGCCGTCTATTACTGTGC 384


RESULT   6
AI007196
LOCUS       AI007196       294 bp    mRNA            EST       12-JUN-1998
DEFINITION  ua73g04.r1 Soares 2NbMT Mus musculus cDNA clone 1363158 5' similar
            to gb:M18512 IG HEAVY CHAIN PRECURSOR V-I REGION (HUMAN); gb:M12376
            Mouse immunoglobulin H-chain V-region pseudogene mRNA, complete
            (MOUSE);, mRNA sequence.
ACCESSION   AI007196
NID         g3216753
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 294)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:896378
            Trace considered overall poor quality
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .294
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library went through two
                     rounds of normalization, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
                     /db_xref="taxon:10090"
                     /clone="1363158"
                     /clone_lib="Soares 2NbMT"
                     /sex="male"
                     /tissue_type="Thymus"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
BASE COUNT       85 a     68 c     77 g     64 t
ORIGIN


  Query Match             18.6%;  Score 133.8;  DB 28;  Length 294;
  Best Local Similarity   70.9%;  Pred. No. 1.2e-31;
  Matches  205;  Conservative    0;  Mismatches   82;  Indels    2;  Gaps    2;
```

```
Qy    358 GTGCAGCTGCAGCAGTCTGGGGTTGAGC-TTGTCCGAGGAGGGGCCTTAGTCAAGTTGTC 416
          || || ||||||||| |||||| ||||| | ||  |    ||| | | ||| ||| |||
Db      2 GTCCAACTGCAGCAGCCTGGGGCTGAGCATGGTGAGGACTGGGTCTTCAGTGAAGCTGTA 61

Qy    417 CTGCAAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCC 476
          |||||| | ||| | || || | | |    ||| |  || || ||||||||||||||||
Db     62 CTGCAAGGATTCAGGCTACACCTTCACCAGCTACTGGATGCATTGGGTGAAGCAGAGGCA 121

Qy    477 TGAACAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATGA 536
          |  ||| ||| | |||||||||||    ||| | |||    || ||   |||   ||  |
Db    122 TATACAAGGCATTGAATGGATTGGTAACATTGACCCTTCAGATAGTGGAACTCACTACAA 181

Qy    537 CCCGAAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTAC-C 595
           |  || ||| ||| ||||||||   | || | ||||| || | ||| | | ||||||
Db    182 TCAAAAGTTCAAGGACAAGGCCACATTGACTGTAGACAAATTCACCAGCACAGCCTACAG 241

Qy    596 TTCAGCTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGC 644
          | ||||||||||| ||||||||||||||| |||| ||||||||||||||
Db    242 TGCAGCTCAGCAGACTGACATCTGAGGACTCTGCGGTCTATTACTGTGC 290


RESULT    7
AA456778
LOCUS       AA456778      400 bp    mRNA            EST       06-JUN-1997
DEFINITION  zw27f11.r1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 770541
            5' similar to gb:X06764 IG KAPPA CHAIN PRECURSOR V-III REGION
            (HUMAN);, mRNA sequence.
ACCESSION   AA456778
NID         g2177199
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 400)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Trace considered overall poor quality
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .400
                     /organism="Homo sapiens"
                     /note="Organ: ovary; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library constructed by Bento Soares and
                     M.Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="770541"
                     /clone_lib="Soares ovary tumor NbHOT"
                     /sex="Female"
                     /tissue_type="ovarian tumor"
```

```
                     /lab_host="DH10B (ampicillin resistant)"
     mRNA            <1. .>400
                     /db_xref="GDB:5980414"
BASE COUNT       90 a    120 c    106 g     84 t
ORIGIN


  Query Match            17.1%;  Score 122.8;  DB 20;  Length 400;
  Best Local Similarity   65.2%;  Pred. No. 3.7e-28;
  Matches  214;  Conservative    0;  Mismatches  107;  Indels    7;  Gaps    2;

Qy      1 ATGCAGCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATG 60
          ||   | |||| ||||||||||  | | |||||   |||||||||||||  | |||| |
Db     61 ATTGTGTTGACGCAGTCTCCAGGCACCCTGTCTTTGTCTCCAGGGGAAAGAGCCACCCTC 120

Qy     61 ACCTGCAGTGCCAGCTCAAGTGTAAGTAACA------TGCACTGGTATCAGCAGAAGTCA 114
           ||||||| |||||    ||||| || | ||      |   |||||| ||||||||  |
Db    121 TCCTGCAGGGCCAGTCACAGTGTTAGCAGCAGCTGCTTAGCCTGGTACCAGCAGAAACCT 180

Qy    115 AGCACCTCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGT 174
           ||    | ||||  |||    | ||||   ||||||    |||  ||||  ||||||
Db    181 GGCCAGGCTCCCAGGCTCCTCATCTATGGTGCATCCAGCAGGGCAACTGGCATCCCAGAC 240

Qy    175 CGCTTCAGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGC-AGCATGGAGGC 233
           | |||||||||||||||||||  |     | | ||||||  ||| || ||  ||||| |
Db    241 AGGTTCAGTGGCAGTGGGTCTGAGACAGACTTCACTCTCAACATCGGCAAGACTGGAGCC 300

Qy    234 TGAAGATGCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTC 293
          ||| |||  |||    ||||| ||| | |||     | | |  || |   ||||||||
Db    301 TGACGATTTTGCAGTGTATTACTGTCAGCAGTATGGTAGCTCACCGTGGACGTTCGGCCA 360

Qy    294 GGGGACAAAGTTGGAAATAAAAGGTTCT 321
           |||||  || ||||||| ||| |  ||
Db    361 AGGGACCCAGGTGGAAATCAAACGAACT 388


RESULT    8
T27593
LOCUS       T27593        393 bp    mRNA            EST       06-SEP-1995
DEFINITION  EST100653 Homo sapiens cDNA 5' end similar to immunoglobulin kappa
            light chain, V region (GB:L01279) (HT:3043).
ACCESSION   T27593
NID         g609691
KEYWORDS    EST.
SOURCE      human primer=M13 Reverse library=Human Pancreas.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 393)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.-W.,
            Clayton,R.A., Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            FitzGerald,L.M., FitzHugh,W.M., Fritchman,J.L., Geoghagen,N.S.M.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle Jr,P.S.,
            Kelley,J.M., Klimek,K.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M.,
            Merrick,J.M., Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T.,
            Pellegrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L.,
            Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R.,
            Weidman,J.F., Li,Y., Bednarik,D.P., Cao,L., Cepeda,M.A.,
            Coleman,T.A., Collins,E.-J., Dimke,D., Feng,P., Ferrie,A.,
            Fischer,C., Hastings,G.A., He,W.-W., Hu,J.-S., Greene,J.M.,
            Gruber,J., Hudson,P., Kim,A., Kozak,D.L., Kunsch,C., Ji,H., Li,H.,
            Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.-F., Wing,J., Xu,C.,
            Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A.,
            Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
  TITLE     Initial Assessment of Human Gene Diversity and Expression Patterns
            Based Upon 52 Million Basepairs of cDNA Sequence
  JOURNAL   Unpublished (1995)
COMMENT     Other_ESTs: THC24356
            Contact: Venter, JC
            The Institute for Genomic Research
            932 Clopper Rd, Gaithersburg, MD 20878
            Tel: 3018699056
            Fax: 3018699423
            Email: tdbinfo@tdb.tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please contact the TIGR Database
            (tdbinfo@tdb.tigr.org).
FEATURES             Location/Qualifiers
     source          1. .393
                     /organism="Homo sapiens"
     mRNA            <1. .>393
BASE COUNT       88 a    113 c     96 g     94 t      2 others
ORIGIN


  Query Match             17.0%;  Score 122.2;  DB 11;  Length 393;
  Best Local Similarity   64.9%;  Pred. No. 5.7e-28;
  Matches  211;  Conservative    0;  Mismatches  110;  Indels    4;  Gaps    2;

Qy        1 ATGCAGCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATG 60
            || ||| |||||||||||||| |   | |||||||||||  ||| || |  |||||| |
Db       56 ATCCAGATGACCCAGTCTCCATCCTCCCTGTCTGCATCTGTAGGAGACAGAGTCACCCTC 115

Qy       61 ACCTGCAGTGCCAGCTCAAGTGTAAGTAACATGC---ACTGGTATCAGCAGAAGTCAAGC 117
            || ||| | || ||    || | || | | |     | ||||||||||||||  || |
Db      116 ACTTGCCGGGCAAGTCAGAGCATTAGCACCTTTTTAAATTGGTATCAGCAGAAACCAGGG 175

Qy      118 ACCTCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGC 177
            ||  |||| || |||  | | | ||   |||||||  ||     ||| ||||||    |
Db      176 ACAGCCCCTAAGCTCCTGATCTCTGCTGCATCCAATTTGCGAAGTGGGGTCCCATCAAGG 235

Qy      178 TTCAGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAA 237
            |||||||||  ||| ||||| |    || | ||||||| ||||||||  || |  |||||
Db      236 TTCAGTGGCTCTGGATCTGGGACAGATTTCACTCTCACCATCAGCAGTCTGCAACCTGAA 295

Qy      238 GATGCTGCCACTTATTATTGTTATCAGGGGAGTGGGTA-CCCATTCACGTTCGGCTCGGG 296
            |||| ||| ||||| || ||| | ||| |       ||   |    ||| | |||   ||
Db      296 GATGTTGCAACTTACTACTGTCAACAGAGTTNCACTTATGNCTCGGACGGTTGGCCAAGG 355

Qy      297 GACAAAGTTGGAAATAAAAGGTTCT 321
            ||| ||| ||||  | ||| |  ||
Db      356 GACCAAGGTGGAGGTCAAACGGACT 380


RESULT   9
AA295311
LOCUS       AA295311      328 bp    mRNA            EST       18-APR-1997
DEFINITION  EST100471 Pancreas tumor I Homo sapiens cDNA 5' end similar to
            immunoglobulin kappa light chain, VJ regions, mRNA sequence.
ACCESSION   AA295311
NID         g1947646
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 328)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
```

```
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: THC169106
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .328
                     /organism="Homo sapiens"
                     /note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI"
                     /db_xref="ATCC (inhost):190766"
                     /db_xref="taxon:9606"
                     /clone_lib="Pancreas tumor I"
                     /dev_stage="adult"
     mRNA            <1. .>328
BASE COUNT       73 a     93 c     78 g     77 t      7 others
ORIGIN


  Query Match             16.5%;  Score 119;  DB 17;  Length 328;
  Best Local Similarity   65.0%;  Pred. No. 5.4e-27;
  Matches  186;  Conservative    0;  Mismatches   97;  Indels    3;  Gaps    1;

Qy        1 ATGCAGCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATG 60
            ||   | |||| |||  | |||| | | ||||||  | |||||||||||  | |||| |
Db       43 ATAGTGATGACGCAGNTTNCAGCCACCCTGTCTGTGTNTCCAGGGGAAAGAGCCACCCTN 102

Qy       61 ACCTGCAGTGCCAGCTCAAGTGTA---AGTAACATGCACTGGTATCAGCAGAAGTCAAGC 117
             ||||||| |||||    |||||    || ||| |   |||||| ||||||||  |  ||
Db      103 TCCTGCAGGGCCAGTCAGAGTGTTAGCAGCAACTTAGCCTGGTACCAGCAGAAACCTGGC 162

Qy      118 ACCTCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGC 177
                | ||||  |||    | ||||   ||||||    ||   ||||  ||||||   |
Db      163 CAGGCTCCCAGGCTCCTCATCTATGGTGCATCCATCAGGGNCACTGGCATCCCAGCCAGG 222

Qy      178 TTCAGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAA 237
            |||||||||||||||||||| |     | | ||||||  ||||||| | || || |||||
Db      223 TTCAGTGGCAGTGGGTCTGGGACAGAGTTCACTCTCANCATCAGCATCCTGCAGTCTGAA 282

Qy      238 GATGCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCA 283
            |||  | |   |||||| ||| | |||   | |   |   ||||||
Db      283 GATTTTNCAGTTTATTACTGTCAGCAGTATAATAACTGGGCATTCA 328


RESULT  10
AA710291
LOCUS       AA710291      345 bp    mRNA            EST       24-DEC-1997
DEFINITION  vt53a04.r1 Barstead mouse irradiated colon MPLRB7 Mus musculus cDNA
            clone 1166766 5' similar to gb:X02484_rna1 IG KAPPA CHAIN PRECURSOR
            V-IV REGION (HUMAN); gb:J00560 mouse ig kappa mrna from mopc21 &
            other myeloma mrna 3' (MOUSE);, mRNA sequence.
ACCESSION   AA710291
NID         g2720209
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 345)
```

```
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:632678
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 140.
FEATURES             Location/Qualifiers
     source          1. .345
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; Tissue obtained
                     from 8 week old mouse.  Colon was harvested 72 hours after
                     irradiation with 1400 Gys.  1st strand cDNA was primed
                     with  a Not I - oligo(dT) primer
                     [5'TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors [AATTCGGATCCTTG], digested with Not I and cloned
                     into the Not I and Eco RI sites of the modified pT7T3
                     vector. Library constructed by Bob Barstead."
                     /db_xref="taxon:10090"
                     /clone="1166766"
                     /clone_lib="Barstead mouse irradiated colon MPLRB7"
                     /dev_stage="8 weeks"
                     /lab_host="DH10B"
BASE COUNT       95 a     81 c     88 g     81 t
ORIGIN


  Query Match             16.5%;  Score 119;  DB 25;  Length 345;
  Best Local Similarity   63.1%;  Pred. No. 5.5e-27;
  Matches  200;  Conservative    0;  Mismatches  115;  Indels    2;  Gaps    1;

Qy      1 ATGCAGCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATG 60
          ||   | ||||||||||||| |  | |||||||  ||||  | || || | ||||| | | 
Db     28 ATTGTGATGACCCAGTCTCAAAAATTCATGTCCACATCAGCTGGAGACAGGGTCAGCGTC 87

Qy     61 ACCTGCAGTGCCAGCTCAAGT--GTAAGTAACATGCACTGGTATCAGCAGAAGTCAAGCA 118
          |||||||   |   |  ||    |  | ||  |   |||||||||  ||||  || |  
Db     88 ACCTGCAAGCCAGTCAGAATGTGGGTACTAATGTAGCCTGGTATCAAGAGAAACCAGGGC 147

Qy    119 CCTCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCT 178
            || || | |     | ||||     |||||| | | |     ||||||||| | |||||
Db    148 AATCTCCTAGAGCACTGATTTACTCGGCATCCTACCGGTACAGTGGAGTCCCTGATCGCT 207

Qy    179 TCAGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAG 238
          |||  |||||||| ||||| |    || | ||||||| ||||||||   ||  || ||||||
Db    208 TCACAGGCAGTGGATCTGGGACAGATTTCACTCTCACCATCAGCAATGTGCAGTCTGAAG 267

Qy    239 ATGCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGGA 298
          |    ||    ||||  ||| | ||    |   | |  || | |||||||||   |||||
Db    268 ACTTGGCAGAGTATTTCTGTCAGCAATATAACAGCTTTCCGTACACGTTCGGAGGGGGGA 327

Qy    299 CAAAGTTGGAAATAAAA 315
          | ||| |||||||||||
Db    328 CCAAGCTGGAAATAAAA 344


RESULT  11
```

```
AA345486
LOCUS       AA345486       403 bp    mRNA             EST       21-APR-1997
DEFINITION  EST51505 Gall bladder II Homo sapiens cDNA 5' end similar to
            similar to immunoglobulin kappa light chain, V region (GB:X06763),
            mRNA sequence.
ACCESSION   AA345486
NID         g1997722
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 403)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .403
                     /organism="Homo sapiens"
                     /note="Organ: gall bladder; Vector: pBluescript SK-;
                     Site_1: EcoRI; Site_2: XhoI"
                     /db_xref="ATCC (inhost):147008"
                     /db_xref="taxon:9606"
                     /clone_lib="Gall bladder II"
                     /sex="female"
                     /dev_stage="adult, 25 yrs"
     mRNA            <1. .>403
BASE COUNT       86 a    113 c    107 g     93 t      4 others
ORIGIN


  Query Match             16.5%;  Score 118.8;  DB 18;  Length 403;
  Best Local Similarity   66.0%;  Pred. No. 6.7e-27;
  Matches  190;  Conservative    0;  Mismatches   92;  Indels    6;  Gaps    1;

Qy     40 CCAGGGGAAAAGGTCACCATGACCTGCAGTGCCAGCTCAAGTGTAAGTAACATGCA---- 96
          ||||||||||  | | || |  |||||||  |||||     |||| || | ||  ||    
Db     42 CCAGGGGAAAGAGCCGCCCTCTCCTGCAGGGCCAGTCAGCGTGTTAGCAGCAGTCACTTA 101

Qy     96 --CTGGTATCAGCAGAAGTCAAGCACCTCCCCCAAACTCTGGGTTTATGACACATCCAAA 153
            |||||||||||||||  |  ||    | ||||  |||    | ||||   | ||||  
Db    102 GCCTGGTATCAGCAGAAACCTGGCCAGGCTCCCAGGCTCCTCATCTATGGTGCGTCCAGG 161
```

```
Qy    154 CTGGCTTCTGGAGTCCCAGGTCGCTTCAGTGGCAGTGGGTCTGGAAACTCTTACTCTCTC 213
            |||  ||||  ||||||   | |||||||||||||||||||| |     | | |||||
Db    162 AGGGCCACTGGCATCCCAGACAGGTTCAGTGGCAGTGGGTCTGGGACAGACTTCACTCTC 221

Qy    214 ACGATCAGCAGCATGGAGGCTGAAGATGCTGCCACTTATTATTGTTATCAGGGGAGTGGG 273
          || ||||||||  ||||| ||||||||  |||    ||||| ||| | |||    |||
Db    222 ACCATCAGCAGACTGGAGCCTGAAGATTTTGCAGTGTATTACTGTCAGCAGTATGGTGAC 281

Qy    274 TACCCATTCACGTTCGGCTCGGGGACAAAGTTGGAAATAAAAGGTTCT 321
          |  ||    |||||||||   ||||| ||| ||||||| ||| |  ||
Db    282 TCACCTAGGACGTTCGGCCAAGGGACCAAGGTGGAAATCAAACGAACT 329


RESULT  12
AA295093
LOCUS       AA295093     370 bp    mRNA            EST       18-APR-1997
DEFINITION  EST100400 Pancreas tumor I Homo sapiens cDNA 5' end similar to
            similar to immunoglobulin kappa, variable region (GB:Y00640), mRNA
            sequence.
ACCESSION   AA295093
NID         g1947582
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 370)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: THC168243
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .370
                     /organism="Homo sapiens"
                     /note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI"
                     /db_xref="ATCC (inhost):190705"
                     /db_xref="taxon:9606"
                     /clone_lib="Pancreas tumor I"
                     /dev_stage="adult"
     mRNA            <1. .>370
BASE COUNT       85 a    107 c     94 g     77 t      7 others
```

```
ORIGIN

  Query Match              16.3%;  Score 117.6;  DB 17;  Length 370;
  Best Local Similarity    66.9%;  Pred. No. 1.5e-26;
  Matches  178;  Conservative    0;  Mismatches   85;  Indels    3;  Gaps    1;

Qy      1 ATGCAGCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATG 60
          ||   | |||| ||    ||||| | | |||| |  | |||||||||||  | |||| |
Db     50 ATAGTGATGACGCANNTNCCAGCCACCCTGTCCGTGTNTCCAGGGGAAAGAGCCACCCTC 109

Qy     61 ACCTGCAGTGCCAGCTCAAGTGTAAGT---AACATGCACTGGTATCAGCAGAAGTCAAGC 117
           ||||||| |||||    ||||| ||    ||| |   |||||| ||||||||  |  ||
Db    110 TCCTGCAGGGCCAGTCAGAGTGTTAGCAGCAACTTAGCCTGGTACCAGCAGAAACCTGGC 169

Qy    118 ACCTCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGC 177
              | ||||  |||    | ||| |  ||||||    |||  ||||  ||||||   |
Db    170 CAGGCACCCAGGCTCCTNATCTATNATGCATCCACCAGGGCCACTGGTATCCCAGCCAGG 229

Qy    178 TTCAGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAA 237
          |||||||||||||||||||| |     | | ||||||| ||||||||| |  || |||||
Db    230 TTCAGTGGCAGTGGGTCTGGGACAGAGTTCACTCTCACCATCAGCAGCCTACAGTCTGAA 289

Qy    238 GATGCTGCCACTTATTATTGTTATCA 263
          ||| ||||   |||||| ||| | ||
Db    290 GATTCTGCAGTTTATTACTGTCAGCA 315


RESULT  13
AA300732
LOCUS       AA300732     333 bp    mRNA            EST       18-APR-1997
DEFINITION  EST13847 Testis tumor Homo sapiens cDNA 5' end similar to similar
            to immunoglobulin kappa light chain (GB:S49006), mRNA sequence.
ACCESSION   AA300732
NID         g1953300
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 333)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: THC87411
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
```

```
                 Index (http://www.tigr.org/tdb/hgi/hgi.html)
                 Seq primer: M13 Reverse.
FEATURES                  Location/Qualifiers
     source               1. .333
                          /organism="Homo sapiens"
                          /note="Organ: testis; Vector: pBluescript SK-; Site_1:
                          EcoRI; Site_2: XhoI"
                          /db_xref="ATCC (inhost):192041"
                          /db_xref="taxon:9606"
                          /clone_lib="Testis tumor"
                          /sex="male"
                          /dev_stage="adult"
     mRNA                 <1. .>333
BASE COUNT       81 a     85 c     80 g     82 t       5 others
ORIGIN


  Query Match             16.3%;  Score 117.6;  DB 17;  Length 333;
  Best Local Similarity   68.4%;  Pred. No. 1.5e-26;
  Matches  162;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy       96 CTGGTATCAGCAGAAGTCAAGCACCTCCCCCAAACTCTGGGTTTATGACACATCCAAACT 155
            |||||||||||||||  || | |   |||| || |||  | | ||||   ||||||   |
Db        3 CTGGTATCAGCAGAAACCAGGGAAAGCCCCTAAGCTCCTGATCTATGCTGCATCCAGTTT 62

Qy      156 GGCTTCTGGAGTCCCAGGTCGCTTCAGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCAC 215
            |     ||| ||||||    | ||||| |||||||| ||||| |    || | |||||||
Db       63 GCAAAGTGGGGTCCCATCAAGGTTCAGCGGCAGTGGATCTGGGACAGATTTCACTCTCAC 122

Qy      216 GATCAGCAGCATGGAGGCTGAAGATGCTGCCACTTATTATTGTTATCAGGGGAGTGGGTA 275
             ||||||||| || || ||||||||  ||| ||||| |||||| | ||||  |   | |
Db      123 CATCAGCAGCCTGCAGCCTGAAGATTTTGCAACTTACTATTGTCAACAGGCTAACAGTTT 182

Qy      276 CCCATTCACGTTCGGCTCGGGGACAAAGTTGGAAATAAAAGGTTCTACCTCCGGATC 332
            ||| |  |||||||||   ||||| ||| ||||||| ||| |  ||    | | | |
Db      183 CCCGTGGACGTTCGGCCAAGGGACCAAGGTGGAAATCAAACGAACTGTGGCTGCANC 239


RESULT  14
AA291381
LOCUS       AA291381      379 bp    mRNA            EST       08-AUG-1997
DEFINITION  zt44g02.r1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 725234
            5' similar to gb:M18512 IG HEAVY CHAIN PRECURSOR V-I REGION
            (HUMAN);, mRNA sequence.
ACCESSION   AA291381
NID         g1939359
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 379)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1576   Std Error: 0.00
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 361.
FEATURES             Location/Qualifiers
     source          1. .379
```

```
                     /organism="Homo sapiens"
                     /note="Organ: ovary; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGTTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library constructed by Bento Soares and
                     M.Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="725234"
                     /clone_lib="Soares ovary tumor NbHOT"
                     /sex="Female"
                     /tissue_type="ovarian tumor"
                     /lab_host="DH10B (ampicillin resistant)"
     mRNA            <1. .>379
BASE COUNT       85 a     82 c    121 g     91 t
ORIGIN


  Query Match             16.2%;  Score 116.6;  DB 21;  Length 379;
  Best Local Similarity   65.1%;  Pred. No. 3.2e-26;
  Matches  188;  Conservative    0;  Mismatches   99;  Indels    2;  Gaps    1;

Qy      356 GTGTGCAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGT 415
            | || ||| ||  |||||||||   |||| |     || |  ||| |  ||| | | | |
Db       35 GGGTTCAGTTGGTGCAGTCTGGACCTGAGGT--GAAGAAGCTGGGGCCCAGTGAGGGTCT 92

Qy      416 CCTGCAAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGC 475
            |||| ||  | ||||  | || | ||      | |  |||| ||||||||  ||||   |
Db       93 CCTGTAAGACCTCTGGTTACACCTTTGGCACTTTTGGTATCAACTGGGTGCGGCAGGCCC 152

Qy      476 CTGAACAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATG 535
            ||| ||| || || || ||||| ||||||||      | | |||||||| ||    | ||
Db      153 CTGGACAAGGGCTTGAGTGGATGGGATGGATCAGCGTTTACAATGGTAAGACAAACTTTG 212

Qy      536 ACCCGAAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTACC 595
                ||  || |||| ||  | |||  | ||  | ||||||||| | ||  | ||||||
Db      213 CAGGGAGTTTTCAGGACAGAGTCAGCTTGACCACTGACACATCCACGAATACAGCCTACA 272

Qy      596 TTCAGCTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGC 644
            |  |||| || | ||||| |||||| ||||| ||| | |||||||||||
Db      273 TGGAGCTGAGGAACCTGAGATCTGACGACACGGCCATATATTACTGTGC 321


RESULT  15
AA515239
LOCUS       AA515239      418 bp    mRNA            EST       19-AUG-1997
DEFINITION  ng69c07.s1 NCI_CGAP_Lip2 Homo sapiens cDNA clone IMAGE:940044
            similar to gb:M63438 IG KAPPA CHAIN PRECURSOR V-III REGION
            (HUMAN);, mRNA sequence.
ACCESSION   AA515239
NID         g2254839
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 418)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 1070   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 413.
FEATURES             Location/Qualifiers
     source          1. .418
                     /organism="Homo sapiens"
                     /note="Vector: pAMP10; mRNA made from liposarcoma, cDNA
                     made by oligo-dT priming. Non- directionally cloned.
                     Size-selected on agarose gel, average insert  size 600 bp.
                     Reference:  Krizman et al. (1996) Cancer Research
                     56:5380-5383."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:940044"
                     /clone_lib="NCI_CGAP_Lip2"
                     /tissue_type="liposarcoma"
                     /lab_host="DH10B"
     mRNA            <1. .>418
BASE COUNT       99 a    127 c    102 g     90 t
ORIGIN


  Query Match             16.0%;  Score 115.4;  DB 21;  Length 418;
  Best Local Similarity   63.9%;  Pred. No. 7.8e-26;
  Matches  209;  Conservative    0;  Mismatches  111;  Indels    7;  Gaps    2;

Qy      1 ATGCAGCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATG 60
          ||   | |||| ||||||||||  | | |||||   |||||||||||||  | |||| |
Db     86 ATTGTGTTGACGCAGTCTCCAGGCACCCTGTCTTTGTCTCCAGGGGAAAGAGCCACCCTC 145

Qy     61 ACCTGCAGTGCCAGCTCAAGTGTAAGTAACATGCA------CTGGTATCAGCAGAAGTCA 114
           | ||||| |||||    ||| | |  | ||  |       |||||||||||| ||  |
Db    146 TCTTGCAGGGCCAGTCAGAGTATCATCAGCACCCCCTTAGCCTGGTATCAGCAAAAACCT 205

Qy    115 AGCACCTCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGT 174
           ||    | ||||  |||    | || |   |||| ||   |||  ||||| ||||||
Db    206 GGCCAGGCTCCCAGGCTCCTCATCTACGGTGCATCGAACAGGGCCACTGGA-TCCCAGAC 264

Qy    175 CGCTTCAGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCT 234
           | |||||||||| |  |||||| |     | |  |||||| |||  |||  ||||| ||
Db    265 AGGTTCAGTGGCAATTCGTCTGGGACAGACTTCAGTCTCACCATCGCCAGACTGGAGCCT 324

Qy    235 GAAGATGCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCG 294
          |||||| ||||   ||| |||||| | ||     ||   |    ||  || ||||||  |
Db    325 GAAGATTCTGCAGTTTACTATTGTCAGCACTATGGTACCTCATTATGGACATTCGGCCAG 384

Qy    295 GGGACAAAGTTGGAAATAAAAGGTTCT 321
          ||||| ||| ||||||| ||| |  ||
Db    385 GGGACCAAGGTGGAAATCAAACGAACT 411


Search completed: May 13, 1999, 09:40:44
Job time: 362 sec
```

```
                    GenCore version 4.5
          Copyright (c) 1993 - 1998  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         May 13, 1999, 09:55:24 ; Search time 59.56 Seconds
                                         (without alignments)
                                         2274.252 Million cell updates/sec

Title:          US-08-704-178-2
Perfect score:  720
Sequence:       1 ATGCAGCTGACCCAGTCTCC.........CGGTCACCGTCTCCTCATAA 720

Scoring table:  IDENTITY_NUC

Searched:       240622 seqs, 94065609 residues

Database :      N_Geneseq_34:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                              SUMMARIES
                     %
Result             Query
  No.     Score    Match Length DB  ID                    Description
 ----------------------------------------------------------------------------
    1       720    100.0    720  1  Q55181                Sequence encoding
    2       720    100.0    720  1  T65007                Single-chain anti-
    3     433.4     60.2    711  1  T17728                Anti-erbB2 scFv cD
    4       433     60.1    711  1  T65006                Single-chain anti-
    5       432     60.0    732  1  Q73678                Fv(GP-2) immunosup
    6     429.8     59.7    711  1  Q55180                Sequence encoding
    7     404.4     56.2    738  1  Q12405                OVB3 light and hea
    8     383.8     53.3    720  1  T91615                cDNA encoding an a
    9     326.4     45.3    735  1  T91614                cDNA encoding an a
   10       322     44.7    720  1  Q51540                Coding sequence of
   11       322     44.7    720  1  T13739                Single chain bindi
   12       322     44.7    720  1  T36463                18-2-3/TRY202' sin
   13     321.6     44.7    724  1  Q05714                18-2-3-/TRY202'. S
   14     321.2     44.6    741  1  Q51538                Coding sequence of
   15     321.2     44.6    741  1  T13737                Single chain bindi
   16     321.2     44.6    741  1  T36461                TRY59 single chain
   17     319.6     44.4    741  1  Q05712                TRY59. Single poly
   18     317.2     44.1    753  1  Q05713                TRY104b. Single po
   19     317.2     44.1    753  1  Q51539                Coding sequence of
   20     317.2     44.1    753  1  T13738                Single chain bindi
   21     317.2     44.1    753  1  T36462                TRY104b single cha
   22     312.6     43.4    906  1  V00611                Anti-human SC sing
   23     310.2     43.1    729  1  Q51541                Coding sequence of
   24     310.2     43.1    729  1  T13740                Single chain bindi
   25     310.2     43.1    729  1  T36464                18-2-3/TRY59 singl
   26       307     42.6    729  1  Q05715                18-2-3-/TRY59. Sin
   27     300.8     41.8   6727  1  V63616                Nucleic acid seque
   28     299.6     41.6   1320  1  V63617                Nucleic acid seque
   29     299.6     41.6   6799  1  V63619                Nucleic acid seque
   30     299.6     41.6   1299  1  V63620                Nucleic acid seque
   31       299     41.5    729  1  Q73679                Fv(GP-4) immunosup
   32     298.6     41.5   1679  1  T58130                Single chain anti-
   33     298.6     41.5   1679  1  V08176                H22-anti-CEA antib
   34     293.6     40.8   1460  1  Q43291                Sequence encoding
   35     293.6     40.8    803  1  T29059                PLAP CC49/212 SCA
   36     292.6     40.6    797  1  T29058                CC49/212 SCA PLAP
   37     290.6     40.4    731  1  Q43287                Sequence encoding
   38       289     40.1    819  1  Q20381                Sequence encoding
   39     288.8     40.1    861  1  Q11098                Encodes gamma heav
   40     287.8     40.0    711  1  Q05711                TRY61. Single poly
   41     287.8     40.0    711  1  Q51537                Coding sequence of
   42     287.8     40.0    711  1  T13736                Single chain bindi
   43     287.8     40.0    711  1  T36460                TRY61 single chain
```

```
   44     285.4     39.6   1065  1  T94963                R. pipiens recombi
   45     285.4     39.6   1137  1  T94964                R. pipiens recombi

                              ALIGNMENTS

RESULT   1
Q55181
ID   Q55181 standard; cDNA; 720 BP.
AC   Q55181;
DT   21-JUL-1994  (first entry)
DE   Sequence encoding the single chain anti-erbB2 antibody, Ab no.21.
KW   Single chain anti-erbB1 antibody; cancer therapy; prevention;
KW   monoclonal antibody; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   cds             1. .171
FT                   /*tag=  a
PN   WO9400136-A.
PD   06-JAN-1994.
PF   21-OCT-1992; U08545.
PR   30-JUN-1992; US-906555.
PA   (MOLE-) MOLECULAR ONCOLOGY INC.
PI   Kasprzyk PG,  King CR;
DR   WPI; 94-025878/03.
DR   P-PSDB; R45443.
PT   Treatment of malignancies over-expressing ERB-[2 - using at least
PT   2 monoclonal antibodies which recognise different epitopes on
PT   gp185
PS   Example; Fig 8; 37pp; English.
CC   The source of human erbB-2 protein for the prodn. of antibodies no.
CC   23 and 21 (Q55180/R45442; Q55181/R45443) is a NIH/3T3 cell
CC   engineered to express the human erbB-2 protein on its surface (N/
CC   erbB-2). Abs no. 21 and 23 are directed against the extracellular
CC   domain of gp185 erbB-2. Nude mice manipulated to produce rapidly
CC   growing tumours were used in a trial of the efficacy of the Abs. In
CC   animals given a combination of the 2 Abs, tumours completely
CC   regressed after 11 days.
SQ   Sequence   720 BP;    184 A;    178 C;    179 G;    179 T;


  Query Match              100.0%;  Score 720;  DB 1;  Length 720;
  Best Local Similarity    100.0%;  Pred. No. 2e-182;
  Matches  720;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATGCAGCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGCAGCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATG 60

Qy     61 ACCTGCAGTGCCAGCTCAAGTGTAAGTAACATGCACTGGTATCAGCAGAAGTCAAGCACC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACCTGCAGTGCCAGCTCAAGTGTAAGTAACATGCACTGGTATCAGCAGAAGTCAAGCACC 120

Qy    121 TCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTTC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTTC 180

Qy    181 AGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGAT 240

Qy    241 GCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGGACA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGGACA 300

Qy    301 AAGTTGGAAATAAAAGGTTCTACCTCCGGATCTGGTAAATCTTCTGAAGGTAAAGGTGTG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AAGTTGGAAATAAAAGGTTCTACCTCCGGATCTGGTAAATCTTCTGAAGGTAAAGGTGTG 360

Qy    361 CAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGTCCTGC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGTCCTGC 420
```

```
Qy    421 AAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCCTGAA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCCTGAA 480

Qy    481 CAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATGACCCG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATGACCCG 540

Qy    541 AAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTACCTTCAG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 AAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTACCTTCAG 600

Qy    601 CTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGCTTCTTATTACTACTAT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGCTTCTTATTACTACTAT 660

Qy    661 AGTGCTTACTATGCTATGTACTACTGGGGTCAAGGAACCTCGGTCACCGTCTCCTCATAA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 AGTGCTTACTATGCTATGTACTACTGGGGTCAAGGAACCTCGGTCACCGTCTCCTCATAA 720


RESULT   2
T65007
ID   T65007 standard; cDNA; 720 BP.
AC   T65007;
DT   05-JUN-1997  (first entry)
DE   Single-chain anti-erbB2 antibody e21(Fv) cDNA.
KW   Single chain antibody; variable region; light chain; heavy chain;
KW   breast cancer; ovarian cancer; non-small cell lung carcinoma;
KW   immunodiagnosis; treatment; cytotoxic agent; erbB-2; ds.
OS   Mus musculus.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   mat_peptide     1. .720
FT                   /*tag= a
FT                   /product= e21(Fv)
PN   US5587458-A.
PD   24-DEC-1996.
PF   07-OCT-1991; 772270.
PR   07-OCT-1991; US-772270.
PR   30-JUN-1992; US-906555.
PR   14-MAY-1993; US-061092.
PA   (ARON-) ARONEX PHARM INC.
PI   Bird RE,  Kasprzyk PG,  King CR;
DR   WPI; 97-064831/06.
DR   P-PSDB; W15186.
PT   Single chain antibodies specific for erbB-2 protein, gp185  - with
PT   labels or cytotoxin, useful for detection and treatment of tumour
PT   cells expressing this protein
PS   Example 9; Columns 27-30; 28pp; English.
CC   The present cDNA sequence codes for a claimed single-chain antibody,
CC   designated e21(Fv), which binds to erbB-2. Monoclonal antibody e21
CC   was generated by immunising mice with N/erbB-2 cells overexpressing
CC   the gp185 protein, removing spleen cells and producing hybridomas
CC   by standard techniques. Messenger RNA coding for the anti-erbB-2
CC   monoclonal antibody was isolated and converted to cDNA. Regions
CC   coding for the heavy- and light- chain variable regions were then
CC   amplified by PCR and joined via a sequence encoding a peptide
CC   linker. The resulting single-chain antibody is useful for in vitro
CC   diagnosis of tumour cells which overexpress the erbB-2 gp185
CC   marker, e.g. breast, ovarian and non-small cell lung carcinomas,
CC   and, when coupled to a cytotoxic agent, to treat such tumours.
SQ   Sequence   720 BP;    184 A;    178 C;    179 G;    179 T;


  Query Match             100.0%;  Score 720;  DB 1;  Length 720;
  Best Local Similarity   100.0%;  Pred. No. 2e-182;
  Matches  720;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATGCAGCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGCAGCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATG 60

Qy     61 ACCTGCAGTGCCAGCTCAAGTGTAAGTAACATGCACTGGTATCAGCAGAAGTCAAGCACC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACCTGCAGTGCCAGCTCAAGTGTAAGTAACATGCACTGGTATCAGCAGAAGTCAAGCACC 120

Qy    121 TCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTTC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTTC 180

Qy    181 AGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGAT 240

Qy    241 GCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGGACA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGGACA 300

Qy    301 AAGTTGGAAATAAAAGGTTCTACCTCCGGATCTGGTAAATCTTCTGAAGGTAAAGGTGTG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AAGTTGGAAATAAAAGGTTCTACCTCCGGATCTGGTAAATCTTCTGAAGGTAAAGGTGTG 360

Qy    361 CAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGTCCTGC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGTCCTGC 420

Qy    421 AAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCCTGAA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCCTGAA 480

Qy    481 CAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATGACCCG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATGACCCG 540

Qy    541 AAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTACCTTCAG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 AAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTACCTTCAG 600

Qy    601 CTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGCTTCTTATTACTACTAT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGCTTCTTATTACTACTAT 660

Qy    661 AGTGCTTACTATGCTATGTACTACTGGGGTCAAGGAACCTCGGTCACCGTCTCCTCATAA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 AGTGCTTACTATGCTATGTACTACTGGGGTCAAGGAACCTCGGTCACCGTCTCCTCATAA 720


RESULT   3
T17728
ID   T17728 standard; cDNA; 711 BP.
AC   T17728;
DT   21-MAY-1996  (first entry)
DE   Anti-erbB2 scFv cDNA.
KW   Oncoprotein; erbB2; cell proliferation; tumour; cancer;
KW   intracellular antibody homologue; single chain antibody; scFv;
KW   gene therapy; ds.
OS   Synthetic.
PN   WO9607321-A1.
PD   14-MAR-1996.
PF   23-AUG-1995; U10740.
PR   06-SEP-1994; US-301339.
PR   06-JUN-1995; US-468252.
PA   (UABR-) UAB RES FOUND.
PI   Curiel DT,  Deshane J;
DR   WPI; 96-171307/17.
DR   P-PSDB; R94020.
PT   Inhibition of proliferation or survival of, esp. malignant erbB2,
PT   cells - by introducing nucleic acid mol. encoding antibody homologue
PT   which is expressed and binds, pref. erbB2, protein intracellularly
PS   Claim 42; Page 29-30; 48pp; English.
CC   A nucleic acid comprises a first sequence encoding a signal peptide
CC   (R94019) linked to a second sequence (T17728) encoding a single
CC   chain Fv fragment (R94020) that binds a human erbB2 oncoprotein.
CC   The anti-erbB2 sFv portion is obtained by PCR using e23scFv
CC   plasmid as template.  The signal peptide directs the scFv to the
CC   endoplasmic reticulum.  The nucleic acid is incorporated into a
CC   plasmid or viral vector to facilitate expression of the scFv antibody
CC   homologue within e.g. an epithelial carcinoma cell.  Intracellular
CC   expression of the homologue inhibits surface expression of erbB2 and
CC   thereby inhibits cell proliferation and cell survival and decreases
CC   tumorigenicity.
SQ   Sequence   711 BP;    175 A;    183 C;    189 G;    164 T;


  Query Match             60.2%;  Score 433.4;  DB 1;  Length 711;
  Best Local Similarity   79.6%;  Pred. No. 2.2e-106;
  Matches  512;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy          2 TGCAGCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATGA 61
              |||||||||||||||||||||||||| | |||||||||||||||||| |||||||| ||||
Db          8 TGCAGCTGACCCAGTCTCCAGCAATCCTCTCTGCATCTCCAGGGGAGAAGGTCACAATGA 67

Qy         62 CCTGCAGTGCCAGCTCAAGTGTAAGTAACATGCACTGGTATCAGCAGAAGTCAAGCACCT 121
              | ||||| |||| | |||||||||| ||||||||||||||||||||||||| || |  |||
Db         68 CTTGCAGGGCCACCCCAAGTGTAAGTTACATGCACTGGTATCAGCAGAAGCCAGGATCCT 127

Qy        122 CCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTTCA 181
              |||||||||  ||| |||||  ||||||||| ||||||||||||||||| | ||||||||
Db        128 CCCCCAAACCTTGGATTTATACCACATCCAACCTGGCTTCTGGAGTCCCTGCTCGCTTCA 187

Qy        182 GTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGATG 241
              ||||| |||||||||| | ||||||||||||||  |||||||  |||||||||||||||
Db        188 GTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGAGGCTGAAGATG 247

Qy        242 CTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGGACAA 301
              ||||||||||||| ||  | ||| ||||| |   ||||  |||||||||   |||| | |
Db        248 CTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGGAGGGGGGTCCA 307

Qy        302 AGTTGGAAATAAAAGGTTCTACCTCCGGATCTGGTAAATCTTCTGAAGGTAAAGGTGTGC 361
              || |||||||||||||||||||||||| || |||||||||||||||||||||||||||||
Db        308 AGCTGGAAATAAAAGGTTCTACCTCTGGTTCTGGTAAATCTTCTGAAGGTAAAGGTGTGC 367

Qy        362 AGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGTCCTGCA 421
              |||||||| |||| ||   |||| | |||      || |  | | | |||  | |||||||
Db        368 AGCTGCAGGAGTCAGGACCTGAGGTGGTCAAGCCTGGAGGTTCAATGAAGATATCCTGCA 427

Qy        422 AAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCCTGAAC 481
              |  ||||||  | |    | |  | | |    ||  |||||||||||||||| | || | 
Db        428 AGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAAGCAGAGCCATGGAA 487

Qy        482 AGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATGACCCGA 541
              ||  ||| || |||||||||   ||| ||||| | |||||| |||||   ||  ||| ||
Db        488 AGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATACTAACTACAACCAGA 547

Qy        542 AATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTACCTTCAGC 601
              | ||| ||||||||||||   | || | |||||  || ||||| | | |||||| |  |||
Db        548 AGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCACAGCCTACATGGAGC 607

Qy        602 TCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGC 644
              ||  ||| ||||||||||||||| |||| |||||||||||||||
Db        608 TCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGC 650


RESULT   4
T65006
ID   T65006 standard; cDNA; 711 BP.
AC   T65006;
DT   05-JUN-1997  (first entry)
DE   Single-chain anti-erbB2 antibody e23(Fv) cDNA.
KW   Single chain antibody; variable region; light chain; heavy chain;
KW   breast cancer; ovarian cancer; non-small cell lung carcinoma;
KW   immunodiagnosis; treatment; cytotoxic agent; erbB-2; ds.
OS   Mus musculus.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   mat_peptide     1. .711
FT                   /*tag=  a
```

```
FT                   /product= e23(Fv)
PN   US5587458-A.
PD   24-DEC-1996.
PF   07-OCT-1991; 772270.
PR   07-OCT-1991; US-772270.
PR   30-JUN-1992; US-906555.
PR   14-MAY-1993; US-061092.
PA   (ARON-) ARONEX PHARM INC.
PI   Bird RE,  Kasprzyk PG,  King CR;
DR   WPI; 97-064831/06.
DR   P-PSDB; W15185.
PT   Single chain antibodies specific for erbB-2 protein, gp185  - with
PT   labels or cytotoxin, useful for detection and treatment of tumour
PT   cells expressing this protein
PS   Example 8; Columns 25-28; 28pp; English.
CC   The present cDNA sequence codes for a claimed single-chain antibody,
CC   designated e23(Fv), which binds to erbB-2. Monoclonal antibody e23
CC   was generated by immunising mice with N/erbB-2 cells overexpressing
CC   the gp185 protein, removing spleen cells and producing hybridomas
CC   by standard techniques. Messenger RNA coding for the anti-erbB-2
CC   monoclonal antibody was isolated and converted to cDNA. Regions
CC   coding for the heavy- and light- chain variable regions were then
CC   amplified by PCR and joined via a sequence encoding a peptide
CC   linker. The resulting single-chain antibody is useful for in vitro
CC   diagnosis of tumour cells which overexpress the erbB-2 gp185
CC   marker, e.g. breast, ovarian and non-small cell lung carcinomas,
CC   and, when coupled to a cytotoxic agent, to treat such tumours.
SQ   Sequence   711 BP;    175 A;    182 C;    189 G;    164 T;


  Query Match             60.1%;  Score 433;  DB 1;  Length 711;
  Best Local Similarity   79.5%;  Pred. No. 2.8e-106;
  Matches  511;  Conservative    1;  Mismatches  131;  Indels    0;  Gaps    0;

Qy          2 TGCAGCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATGA 61
              |||||||||||||||||||||||||| ||||||||||||||||||| |||||||| ||||
Db          8 TGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTCACAATGA 67

Qy         62 CCTGCAGTGCCAGCTCAAGTGTAAGTAACATGCACTGGTATCAGCAGAAGTCAAGCACCT 121
              | ||||| |||| | |||||||||| ||||||||||||||||||||||||| || |  |||
Db         68 CTTGCAGGGCCACCCCAAGTGTAAGTTACATGCACTGGTATCAGCAGAAGCCAGGATCCT 127

Qy        122 CCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTTCA 181
              |||||||||  ||| |||||  ||||||||| ||:|||||||||||||| | ||||||||
Db        128 CCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGTCCCTGCTCGCTTCA 187

Qy        182 GTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGATG 241
              ||||| |||||||||| | ||||||||||||||  |||||||  |||||||||||||||
Db        188 GTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAGTCAGCAGAGTGGAGGCTGAAGATG 247

Qy        242 CTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGGACAA 301
              ||||||||||||| ||  | ||| ||||| |   ||||  |||||||||   |||| | |
Db        248 CTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGGAGGGGGGTCCA 307

Qy        302 AGTTGGAAATAAAAGGTTCTACCTCCGGATCTGGTAAATCTTCTGAAGGTAAAGGTGTGC 361
              || |||||||||||||||||||||||| || |||||||||||||||||||||||||||||
Db        308 AGCTGGAAATAAAAGGTTCTACCTCTGGTTCTGGTAAATCTTCTGAAGGTAAAGGTGTGC 367

Qy        362 AGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGTCCTGCA 421
              |||||||| |||| ||   |||| | |||      || |  | | | |||  | |||||||
Db        368 AGCTGCAGGAGTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTCAATGAAGATATCCTGCA 427

Qy        422 AAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCCTGAAC 481
              |  ||||||  | |    | |  | | |    ||  |||||||||||||||| | || | 
Db        428 AGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAAGCAGAGCCATGGAA 487

Qy        482 AGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATGACCCGA 541
              ||  ||| || |||||||||   ||| ||||| | |||||| |||||   ||  ||| ||
Db        488 AGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATACTAACTACAACCAGA 547

Qy        542 AATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTACCTTCAGC 601
              | ||| ||||||||||||   | || | |||||  || ||||| | | |||||| |  |||
Db        548 AGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCACAGCCTACATGGAGC 607
```

```
Qy    602 TCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGC 644
          ||  ||| ||||||||||||||| |||| ||||||||||||||
Db    608 TCCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGC 650


RESULT   5
Q73678
ID   Q73678 standard; DNA; 732 BP.
AC   Q73678;
DT   09-MAY-1995  (first entry)
DE   Fv(GP-2) immunosuppressive.
KW   MAb; monoclonal antibody; hybridoma; interleukin-2; IL-2;
KW   Fv; antibody variable region; GP-2; Fv(GP-2);
KW   immunosuppressive; ss.
OS   Mus sp.
PN   EP-621338-A.
PD   26-OCT-1994.
PF   21-APR-1994; 106257.
PR   21-APR-1993; JP-094491.
PR   07-MAR-1994; JP-036065.
PA   (AJIN ) AJINOMOTO KK.
PI   Hamura J,  Kanayama Y,  Nakazawa H,  Shimamura T,  Sugamura K;
PI   Takeshita T;
DR   WPI; 94-325948/41.
DR   P-PSDB; R60780.
PT   Immunosuppressant polypeptide - has ability to block
PT   interleukin-2 response
PS   Claim 17; Page 29; 37pp; English.
CC   MAb capable of binding to the gamma chain of the IL-2 receptor, and
CC   thus of blocking the IL-2 response, is produced by mouse hybridoma
CC   line GP-2 (FERM BP-4641).  DNA encoding the variable region of
CC   this MAb was expressed in E. coli, yielding Fv(GP-2) with
CC   immunosuppressive activity.
SQ   Sequence   732 BP;    180 A;    204 C;    182 G;    166 T;


  Query Match             60.0%;  Score 432;  DB 1;  Length 732;
  Best Local Similarity   76.7%;  Pred. No. 5.2e-106;
  Matches  557;  Conservative    0;  Mismatches  160;  Indels    9;  Gaps    2;

Qy      1 ATGCAGCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATG 60
          || | |||||||||||||||| |||||||||||||||||| |||||||  ||||||||||
Db      7 ATCCTGCTGACCCAGTCTCCATCAATCATGTCTGCATCTCTAGGGGAACGGGTCACCATG 66

Qy     61 ACCTGCAGTGCCAGCTCAAGTGTA------AGTAACATGCACTGGTATCAGCAGAAGTCA 114
          ||||||| ||||||||||||||||      ||| || |||||||||| ||||||||| ||
Db     67 ACCTGCACTGCCAGCTCAAGTGTAAGTTCCAGTTACTTGCACTGGTACCAGCAGAAGCCA 126

Qy    115 AGCACCTCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGT 174
           |  ||||||||||||||||| |||||  ||||||||| ||||||||||||||||||| |
Db    127 GGATCCTCCCCCAAACTCTGGATTTATAGCACATCCAACCTGGCTTCTGGAGTCCCAGCT 186

Qy    175 CGCTTCAGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCT 234
          ||||||||||||||||||||||| | ||||||||||||||| ||||||||||||||||||
Db    187 CGCTTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGCATGGAGGCT 246

Qy    235 GAAGATGCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCG 294
          |||||||||||||||||||| ||  | |||     | | | |||  ||||||||||  |
Db    247 GAAGATGCTGCCACTTATTACTGCCACCAGTATCATCGTTCCCCGCTCACGTTCGGTGCT 306

Qy    295 GGGACAAAGTTGGAAATAAAAGGTTCTA---CCTCCGGATCTGGTAAATCTTCTGAAGGT 351
          ||||| ||| ||||  | ||||     |   |||| ||||||||       ||  || |
Db    307 GGGACCAAGCTGGAGCTCAAAGTCGACAAATCCTCAGGATCTGGCTCCGAATCCAAAAGC 366

Qy    352 AAAGGTGTGCAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAG 411
          |     ||  | ||  || |||||||   |||||| ||  |    || || | ||| |||
Db    367 ACGCAGGTCAAACTCGAGGAGTCTGGATCTGAGCTGGTGAGGCCTGGAGCTTCAGTGAAG 426

Qy    412 TTGTCCTGCAAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAG 471
           |||||||||| ||||||| || ||   | |    ||| |  || |||||||||||||||
Db    427 CTGTCCTGCAAGGCTTCTGGCTACACATTCACCAGCTACTGGATGCACTGGGTGAAGCAG 486

Qy    472 AGGCCTGAACAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTA 531
          |||| || ||| ||||| || |||||||||   ||| ||||||  | ||||| ||||
Db    487 AGGCATGGACAAGGCCTTGAGTGGATTGGAAATATTTATCCTGGTAGTGGTAGTACTAAC 546

Qy    532 TATGACCCGAAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCC 591
          || ||   ||| ||| || |||||||||   | || | |||||||||||||| | | |||
Db    547 TACGATGAGAAGTTCAAGAGCAAGGCCACACTGACTGTAGACACATCCTCCAGCACAGCC 606

Qy    592 TACCTTCAGCTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGCTTCTTAT 651
          ||| | || |||||||||||||||||||||||| |||| |||||||||||| |
Db    607 TACATGCACCTCAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTACAAGAAGC 666

Qy    652 TACTACTATAGTGCTTACTATGCTATGTACTACTGGGGTCAAGGAACCTCGGTCACCGTC 711
            |    |  | |  |||||||||||| |||||||||||||||||||||| |||||||||
Db    667 AGCCGGAACTGGGTCTACTATGCTATGGACTACTGGGGTCAAGGAACCTCAGTCACCGTC 726

Qy    712 TCCTCA 717
          ||||||
Db    727 TCCTCA 732


RESULT   6
Q55180
ID   Q55180 standard; cDNA; 711 BP.
AC   Q55180;
DT   21-JUL-1994  (first entry)
DE   Sequence encoding the single chain anti-erbB2 antibody, Ab no.23.
KW   Single chain anti-erbB1 antibody; cancer therapy; prevention;
KW   monoclonal antibody; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   cds             1. .171
FT                   /*tag=  a
PN   WO9400136-A.
PD   06-JAN-1994.
PF   21-OCT-1992; U08545.
PR   30-JUN-1992; US-906555.
PA   (MOLE-) MOLECULAR ONCOLOGY INC.
PI   Kasprzyk PG,  King CR;
DR   WPI; 94-025878/03.
DR   P-PSDB; R45442.
PT   Treatment of malignancies over-expressing ERB-[2 - using at least
PT   2 monoclonal antibodies which recognise different epitopes on
PT   gp185
PS   Example; Fig 7; 37pp; English.
CC   The source of human erbB-2 protein for the prodn. of antibodies no.
CC   23 and 21 (Q55180/R45442; Q55181/R45443) is a NIH/3T3 cell
CC   engineered to express the human erbB-2 protein on its surface (N/
CC   erbB-2). Abs no. 21 and 23 are directed against the extracellular
CC   domain of gp185 erbB-2. Nude mice manipulated to produce rapidly
CC   growing tumours were used in a trial of the efficacy of the Abs. In
CC   animals given a combination of the 2 Abs, tumours completely
CC   regressed after 11 days.
SQ   Sequence   711 BP;    175 A;    183 C;    187 G;    165 T;


  Query Match             59.7%;  Score 429.8;  DB 1;  Length 711;
  Best Local Similarity   79.2%;  Pred. No. 2e-105;
  Matches  509;  Conservative    1;  Mismatches  133;  Indels    0;  Gaps    0;

Qy      2 TGCAGCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATGA 61
          |||||||||||||||||||||||||| ||||||||||||||||||| |||||||| ||||
Db      8 TGCAGCTGACCCAGTCTCCAGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTCACAATGA 67

Qy     62 CCTGCAGTGCCAGCTCAAGTGTAAGTAACATGCACTGGTATCAGCAGAAGTCAAGCACCT 121
          | ||||| |||| | ||||||||||| ||||||||||||||||||||||| || |  |||
Db     68 CTTGCAGGGCCACCCCAAGTGTAAGTTACATGCACTGGTATCAGCAGAAGCCAGGATCCT 127

Qy    122 CCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTTCA 181
          |||||||||  ||| |||||  ||||||||| ||:|||||||||||||| | ||||||||
Db    128 CCCCCAAACCTTGGATTTATACCACATCCAACCTKGCTTCTGGAGTCCCTGCTCGCTTCA 187

Qy    182 GTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGATG 241
```

```
           ||||| |||||||||| | |||||||||||||||  |||||||  |||||||||||||||
Db     188 GTGGCGGTGGGTCTGGGACCTCTTACTCTCACAGTCAGCAGAGTGGAGGCTGAAGATG 247

Qy     242 CTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGGACAA 301
           ||||||||||||| ||  | ||| ||||| |   ||||  |||||||||   |||| | |
Db     248 CTGCCACTTATTACTGCCAGCAGTGGAGTCGTAGCCCACCCACGTTCGGAGGGGGGTCCA 307

Qy     302 AGTTGGAAATAAAAGGTTCTACCTCCGGATCTGGTAAATCTTCTGAAGGTAAAGGTGTGC 361
           || |||||||||||||||||||||| || |||||||||||||||||||||||||||||||
Db     308 AGCTGGAAATAAAAGGTTCTACCTCTGGTTCTGGTAAATCTTCTGAAGGTAAAGGTGTGC 367

Qy     362 AGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGTCCTGCA 421
           |||||||| | || ||   |||| | ||       || |  | | | ||| | |||||||
Db     368 AGCTGCAGGATTCAGGACCTGAGGTGGTGAAGCCTGGAGGTTCAATGAAGATATCCTGCA 427

Qy     422 AAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCCTGAAC 481
           |  ||||||  | |    | |  | | |    ||  |||||||||||||||| | || |
Db     428 AGACTTCTGGTTACTCATTCACTGGCCACACCATGAACTGGGTGAAGCAGAGCCATGGAA 487

Qy     482 AGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATGACCCGA 541
           ||  ||| || |||||||||   ||| ||||| | |||||| |||||   ||  ||| ||
Db     488 AGAACCTTGAGTGGATTGGACTTATTAATCCTTACAATGGTGATACTAACTACAACCAGA 547

Qy     542 AATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTACCTTCAGC 601
           | ||| ||||||||||||   | || | |||||  || |||| | | |||||| |  |||
Db     548 AGTTCAAGGGCAAGGCCACATTTACTGTAGACAAGTCGTCCAGCACAGCCTACATGGAGC 607

Qy     602 TCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGC 644
           ||  ||| ||||||||||||||| ||||  |||||||||||||
Db     608 TCCTCAGTCTGACATCTGAGGACTCTGCACTCTATTACTGTGC 650

RESULT   7
Q12405
ID   Q12405 standard; DNA; 738 BP.
AC   Q12405;
DT   16-SEP-1991  (first entry)
DE   OVB3 light and heavy chains.
KW   Immunoglobulin; immunotoxin; antibody; variable region; PE40;
KW   pseudomonas exotoxin; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   cds             7. .729
FT                   /*tag= a
FT   misc_rna        7. .327
FT                   /*tag= b
FT                   /label= light chain
FT   misc_rna        328. .369
FT                   /*tag= b
FT                   /label= linker
FT   misc_rna        370. .729
FT                   /*tag= c
FT                   /label= heavy chain
PN   US7463111-A.
PD   28-MAY-1991.
PF   11-JAN-1990; 133066.
PR   11-JAN-1990; US-463111.
PA   (USSH ) NAT INST OF HEALTH.
PI   Pastan I;
DR   WPI; 91-200877/27.
DR   P-PSDB; R12798.
PT   Rapid cloning of antibody genes as single chain immuno-toxins -
PT   by fusing genes with DNA encoding cytotoxic molecules, to kill
PT   target cells
PS   Disclosure; Fig 3; 26pp; English.
CC   The sequence is an NdeI-HindIII fragment encoding the L and H
CC   chains of murine MAb OVB3, joined together via a linker.  The se-
CC   quence is present in clone pOVB3158-1 in which the 3' end is
CC   linked to the 5' end of the coding sequence for PE40, a a Pseudo-
CC   monas exotoxin lacking domain I.  The protein expressed by the
CC   clone can be used to kill specific target cells.
SQ   Sequence   738 BP;    181 A;    193 C;    192 G;    172 T;
```

```
  Query Match             56.2%;  Score 404.4;  DB 1;  Length 738;
  Best Local Similarity   73.7%;  Pred. No. 1.1e-98;
  Matches  529;  Conservative    0;  Mismatches  186;  Indels    3;  Gaps    1;

Qy       6 GCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATGACCTG 65
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15 GCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATGACCTG 74

Qy      66 CAGTGCCAGCTCAAGTGTAAGTAACATGCACTGGTATCAGCAGAAGTCAAGCACCTCCCC 125
           |||||||| ||||| ||||||| ||||||||||||| |||||||||||||||||||||||
Db      75 CAGTGCCATCTCAACTGTAAGTTACATGCACTGGTACCAGCAGAAGTCAAGCACCTCCCC 134

Qy     126 CAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTTCAGTGG 185
           |||||||||| ||||||| |||||||||||||||||||||||||||||||||||||||||
Db     135 CAAACTCTGGATTTATGAGACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTTCAGTGG 194

Qy     186 CAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGATGCTGC 245
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     195 CAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGATGTTGC 254

Qy     246 CACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGGACAAAGTT 305
           ||||||||| |||| |||||||||||||||||| | |||||||||   |||||| ||| |
Db     255 CACTTATTACTGTTTTCAGGGGAGTGGGTACCCGTACACGTTCGGAGGGGGGACCAAGCT 314

Qy     306 GGAAATAAAA---GGTTCTACCTCCGGATCTGGTAAATCTTCTGAAGGTAAAGGTGTGCA 362
           |||  | |||   |||    |||| ||||||||       ||  |||   | | ||||||
Db     315 GGAGCTCAAAGAGGGTAAATCCTCAGGATCTGGCTCCGAATCCAAAGTCGACGATGTGCA 374

Qy     363 GCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGTCCTGCAA 422
           ||||  | |||||||||  |   | || |     || |  |      |  | |||||
Db     375 GCTGGTGGAGTCTGGGGGAGGCTTAGTGCAGCCTGGAGGGTCCCGGCAACTCTCCTGTGC 434

Qy     423 AGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCCTGAACA 482
           ||| ||||  ||||   | |    || |   || ||||||||    |||   || ||  |
Db     435 AGCCTCTGGATTCACTTTCAGTAGCTTTGGAATGCACTGGGTTCGTCAGGCTCCAGAGAA 494

Qy     483 GGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATGACCCGAA 542
           ||| ||||| ||| | | ||  |||  || ||    |  ||  |     ||||     |
Db     495 GGGGCTGGAGTGGGTCGCATACATTAGTCGTGGCGGTAATACCATCTACTATGCAAACAC 554

Qy     543 ATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTACCTTCAGCT 602
           | |  |||||     ||  ||  |   |||||   ||   ||| |    | ||| ||  |
Db     555 AGTGAAGGGCCGATTCACCATCTCCAGAGACAATCCCAAGAACACCCTGTTCCTGCAAAT 614

Qy     603 CAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGCTTCTTATTACTACTATAG 662
            | ||| || |  ||||| ||||| ||| | |||||||||||    |   | |||||  |
Db     615 GACCAGTCTAAGGTCTGACGACACGGCCATGTATTACTGTGCAAGATCCCATTACTACGG 674

Qy     663 TGCTTACTATGCTATGTACTACTGGGGTCAAGGAACCTCGGTCACCGTCTCCTCATAA 720
               | |||||||||| ||||||||||||| |||||| |  |||| ||||||||| ||
Db     675 CTACTTCTATGCTATGGACTACTGGGGTCAGGGAACCACTCTCACAGTCTCCTCAAAA 732

RESULT   8
T91615
ID   T91615 standard; cDNA to mRNA; 720 BP.
AC   T91615;
DT   22-DEC-1997  (first entry)
DE   cDNA encoding an anti-gp130 antibody derived scFv.
KW   Single chain Fv; scFv; antibody; E. coli; anti-T3 antibody;
KW   anti-gp130 antibody; inclusion body; chaperonin; ds.
OS   Synthetic.
PN   J09220092-A.
PD   26-AUG-1997.
PF   15-FEB-1996; 027622.
PR   15-FEB-1996; JP-027622.
PA   (TOYJ ) TOSOH CORP.
DR   WPI; 97-474306/44.
DR   P-PSDB; W25784.
PT   Producing single chain Fv antibody in Escherichia coli - by
PT   expression in an inclusion body, followed by protein folding or by
PT   co-expression with a chaperonin as a soluble fraction
```

```
PS   Claim 4; Page 6-7; 9pp; Japanese.
CC   The sequences given in T91614-15 encode single chain Fv (scFv)
CC   antibodies which are produced in E. coli.  The scFV's are derived
CC   from an anti-T3 antibody or an anti-gp130 antibody and are produced
CC   either by: (1) expression as an inclusion body, followed by folding
CC   (i.e. by denaturation and solubilisation) or (2) expression as a
CC   soluble fraction by co-expression with chaperonin.  Using the methods,
CC   scFv can be produced in E. coli, either in a soluble fraction or in
CC   inclusion bodies.  Like the parent double chain antibodies, the scFv
CC   polypeptides are useful in immunological diagnosis or for separation
CC   and purification, but they lessen the side effects caused by antibody
CC   constant regions.  The availability of scFv polypeptides contributes
CC   to fundamental research and development of diagnostic and therapeutic
CC   drugs.
SQ   Sequence   720 BP;    169 A;    193 C;    190 G;    168 T;


  Query Match             53.3%;  Score 383.8;  DB 1;  Length 720;
  Best Local Similarity   73.0%;  Pred. No. 3.2e-93;
  Matches  522;  Conservative    0;  Mismatches  187;  Indels    6;  Gaps    2;

Qy       6 GCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATGACCTG 65
           ||| ||||||||||||||| | || ||||||||||||||||| |||||||||||||||||
Db       9 GCTCACCCAGTCTCCAGCATTGATATCTGCATCTCCAGGGGAGAAGGTCACCATGACCTG 68

Qy      66 CAGTGCCAGCTCAAGTGTAAGTAACATGCACTGGTATCAGCAGAAGTCAAGCACCTCCCC 125
           || || |||||||||||| | |  |||| | ||||| ||||||||| || |  |||||||
Db      69 CAATGTCAGCTCAAGTGTTACTTCCATGTATTGGTACCAGCAGAAGCCAGGATCCTCCCC 128

Qy     126 CAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTTCAGTGG 185
           ||||| |||| |||||  ||||||||| |||||||||||||||||   ||||| ||||||
Db     129 CAAACCCTGGATTTATCTCACATCCAACCTGGCTTCTGGAGTCCCTCCTCGCTCCAGTGG 188

Qy     186 CAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGATGCTGC 245
           |||||||||||| | | ||||||||||||| ||||||||| |||||||||||||||||||
Db     189 CAGTGGGTCTGGGACCACTTACTCTCTCACAATCAGCAGCTTGGAGGCTGAAGATGCTGC 248

Qy     246 CACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGGACAAAGTT 305
           ||||||||| ||  | ||| |||||    ||||  ||||||||||  | ||||| ||| |
Db     249 CACTTATTACTGCCAGCAGTGGAGTACTAACCCGCTCACGTTCGGTGCTGGGACCAAGCT 308

Qy     306 GGAAATAAAAGGTTCTACCTCCGGATCTGGTAAATCTTCTGAAGGTAAA---GGTGTGCA 362
            ||  |    ||   |  ||| ||   ||||   ||   ||  ||   |   |  || ||
Db     309 CGAGCTGGGTGGCGGTGGCTCGGGCGGTGGTGGGTCGGGTGGCGGCGGATCGGACGTCCA 368

Qy     363 GCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGTCCTGCAA 422
             |||||||||||||   ||| || ||       ||||| | ||| ||| |  |||||||
Db     369 ATTGCAGCAGTCTGGACCTGAACTGGTGAAGCCTGGGGCTTCAGTGAAGATACCCTGCAA 428

Qy     423 AGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCCTGAACA 482
            ||||| |  | ||   | |  |||||  | ||  |||||||||||||||| | || | |
Db     429 GGCTTCAGGATACACATTCACTGACTACAACATGGACTGGGTGAAGCAGAGCCATGGAAA 488

Qy     483 GGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATGACCCGAA 542
           | |||| || |||||||||   ||| || || | | ||||  || | | ||  |||  ||
Db     489 GAGCCTTGAGTGGATTGGAGATATTAATTCTCATAGTGGTGGTATTATCTACAACCAAAA 548

Qy     543 ATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTACCTTCAGCT 602
            ||| ||| ||||||||   | || | ||| |  ||||||| | | |||||| |  ||||
Db     549 GTTCAAGGACAAGGCCATATTGACTGTAGATAAGTCCTCCAGCACAGCCTACATGGAGCT 608

Qy     603 CAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGCTTCTTATTACTACTATAG 662
           | ||||||||||||||||||||||||| |||||||| ||||   |      |||||||
Db     609 CCGCAGCCTGACATCTGAGGACACTGCAGTCTATTATTGTG---CAAGAACCTACTATGT 665

Qy     663 TGCTTACTATGCTATGTACTACTGGGGTCAAGGAACCTCGGTCACCGTCTCCTCA 717
                 | |     |  |||||||||| ||||| ||| |  |||| |||||||||
Db     666 GTACGGCCACTTCTTTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCA 720


RESULT   9
T91614
ID   T91614 standard; cDNA to mRNA; 735 BP.
```

```
AC   T91614;
DT   22-DEC-1997  (first entry)
DE   cDNA encoding an anti-T3 antibody derived scFv.
KW   Single chain Fv; scFv; antibody; E. coli; anti-T3 antibody;
KW   anti-gp130 antibody; inclusion body; chaperonin; ds.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   cds             1. .735
FT                   /*tag=  a
FT                   /transl_except= pos:160. .173, aa:Tyr
FT                   /transl_except= pos:385. .387, aa:Glu
FT                   /note= "No stop codon given"
PN   J09220092-A.
PD   26-AUG-1997.
PF   15-FEB-1996; 027622.
PR   15-FEB-1996; JP-027622.
PA   (TOYJ ) TOSOH CORP.
DR   WPI; 97-474306/44.
DR   P-PSDB; W25783.
PT   Producing single chain Fv antibody in Escherichia coli - by
PT   expression in an inclusion body, followed by protein folding or by
PT   co-expression with a chaperonin as a soluble fraction
PS   Claim 3; Page 5-6; 9pp; Japanese.
CC   The sequences given in T91614-15 encode single chain Fv (scFv)
CC   antibodies which are produced in E.coli.  The scFV's are derived
CC   from an anti-T3 antibody or an anti-gp130 antibody and are produced
CC   either by: (1) expression as an inclusion body, followed by folding
CC   (i.e. by denaturation and solubilisation) or (2) expression as a
CC   soluble fraction by co-expression with chaperonin.  Using the methods,
CC   scFv can be produced in E. coli, either in a soluble fraction or in
CC   inclusion bodies.  Like the parent double chain antibodies, the scFv
CC   polypeptides are useful in immunological diagnosis or for separation
CC   and purification, but they lessen the side effects caused by antibody
CC   constant regions.  The availability of scFv polypeptides contributes
CC   to fundamental research and development of diagnostic and therapeutic
CC   drugs.
SQ   Sequence   735 BP;    165 A;    192 C;    197 G;    181 T;


  Query Match             45.3%;  Score 326.4;  DB 1;  Length 735;
  Best Local Similarity   67.6%;  Pred. No. 5.5e-78;
  Matches  491;  Conservative    0;  Mismatches  226;  Indels    9;  Gaps    2;

Qy       1 ATGCAGCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATG 60
           || |||||||||||||||||| | | ||||  ||||||||| ||||| ||| |||| || 
Db      10 ATTCAGCTGACCCAGTCTCCAACCACCATGGTTGCATCTCCCGGGGAGAAGATCACTATC 69

Qy      61 ACCTGCAGTGCCAGCTCAAGTGTAAG------TAACATGCACTGGTATCAGCAGAAGTCA 114
           ||||||||||||||||||||| ||||      | || |||| ||||||||||||||| ||
Db      70 ACCTGCAGTGCCAGCTCAAGTATAAGTTCCAATTACTTGCATTGGTATCAGCAGAAGCCA 129

Qy     115 AGCACCTCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGT 174
            |   |||||| ||||||| | |||||   |||||||| ||||||||||| ||||||  |
Db     130 GGATTCTCCCCTAAACTCTTGATTTATAGGACATCCAATCTGGCTTCTGGTGTCCCAACT 189

Qy     175 CGCTTCAGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCT 234
           ||||||||||||||||||||||| | ||||||||||||||| ||  ||| ||||||||||
Db     190 CGCTTCAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATTGGCACCATGGAGGCT 249

Qy     235 GAAGATGCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCG 294
           ||||||| ||||||||| || ||  | ||||| ||| |    ||  ||||||||||  | 
Db     250 GAAGATGTTGCCACTTACTACTGCCAGCAGGGTAGTAGTATACCGCTCACGTTCGGTGCT 309

Qy     295 GGGACAAAGTTGGAAATAAAAGGTTCTACCTCCGGATCTGGTAAATCTTCTGAAGGTAAA 354
           ||||| ||| | || ||    ||   |  ||| ||   ||||   ||   ||  ||   |
Db     310 GGGACCAAGCTCGAGATCGGTGGCGGTGGCTCGGGCGGTGGTGGGTCGGGTGGCGGCGGA 369

Qy     355 ---GGTGTGCAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAG 411
              |  ||  |||||||| |||||||||  |   | ||       || |  |   | || 
Db     370 TCCGAGGTCAAGCTGCAGGAGTCTGGGGGAGGCTTAGTGAAGCTTGGCGGGTCCCTGAAA 429

Qy     412 TTGTCCTGCAAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAG 471
            | |||||  |||| ||||  ||||   | |     ||||| ||    |||||    |||
Db    430 CTCTCCTGTGAAGCCTCTGGATTCACTTTCAGTAGTTATTACATGTCTTGGGTTCGCCAG 489

Qy    472 AGGCCTGAACAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTA 531
          |  || ||  || | ||||| | | | | |   ||| ||  | |   |||||  ||
Db    490 ACTCCAGAGAAGAGGCTGGAGTTGGTCGCAGCCATTAATAGTAATGGTGGTACCACCTAC 549

Qy    532 TATGACCCGAAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCC 591
          |||      |   |  |||||     ||  ||  |   |||||   ||   ||| |
Db    550 TATTCAGACACTGTGAAGGGCCGATTCACCATCTCCAGAGACAATGCCAAGAACACCCTG 609

Qy    592 TACCTTCAGCTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGCTTCTTAT 651
          ||||| ||  | ||||| ||||  ||||||||||| ||| | |||||||||||
Db    610 TACCTGCAAATGAGCAGTCTGAAGTCTGAGGACACAGCCTTGTATTACTGTGCAAGCCCG 669

Qy    652 TACTACTATAGTGCTTACTATGCTATGTACTACTGGGGTCAAGGAACCTCGGTCACCGTC 711
            || ||||  |     |||||  |||||||||||||| ||||| ||| |||||||||||
Db    670 GTCTCCTATTATTACCTCTATGTCATGTACTACTGGGGCCAAGGGACCACGGTCACCGTC 729

Qy    712 TCCTCA 717
          ||||||
Db    730 TCCTCA 735


RESULT  10
Q51540
ID   Q51540 standard; DNA; 720 BP.
AC   Q51540;
DT   23-MAY-1994  (first entry)
DE   Coding sequence of polypeptide with affinity for fluorescein.
KW   Monoclonal antibody; MAb; affinity; binding; antigen; diagnostics;
KW   therapy; imaging; purification; biosensors; ss.
OS   Synthetic.
FH   Key              Location/Qualifiers
FT   misc_difference  619. .621
FT                    /*tag= a
FT                    /transl_except= AAT encodes Isoleucine.
FT   misc_difference  658. .660
FT                    /*tag= b
FT                    /transl_except= GAA encodes Glycine.
PN   US5260203-A.
PD   09-NOV-1993.
PF   02-SEP-1986; 902971.
PR   02-SEP-1986; US-902971.
PR   02-SEP-1987; US-092110.
PR   19-JAN-1989; US-299617.
PR   25-APR-1990; US-512910.
PA   (ENZO-) ENZON INC.
PI   Bird RE,  Hardman K,  Ladner RC;
DR   WPI; 93-367875/46.
DR   P-PSDB; R43679.
PT   Single chain poly:peptide for binding antigen - comprising light
PT   and heavy chain antigen binding portions linked by peptide linker
PS   Example 9; Figure 39; 78pp; English.
CC   The single chain polypeptide is derived from the mature light and
CC   heavy chains of a monoclonal antibody (MAb) and has affinity
CC   for a given antigen (Fluorescein).  It comprises a first
CC   polypeptide comprising the antigen binding portion of of the light
CC   chain variable region of an antibody and a second polypeptide
CC   comprising the antigen binding portion of the heavy chain variable
CC   region of an antibody and at least one peptide linker linking the
CC   first and second polypeptide chains.  The resulting single chain
CC   polypeptide can be used in diagnostics, therapy
CC   (in vivo and in vitro), imaging, purifications and biosensors.
CC   This particular single chain binding molecule was designated
CC   18-2-3/TRY202' and contains one linker peptide.
SQ   Sequence   720 BP;    180 A;    185 C;    175 G;    180 T;


  Query Match            44.7%;  Score 322;  DB 1;  Length 720;
  Best Local Similarity  68.7%;  Pred. No. 8e-77;
  Matches  493;  Conservative    0;  Mismatches  210;  Indels   15;  Gaps    3;

Qy      6 GCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATGACCTG 65
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12 GCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATGACCTG 71

Qy     66 CAGTGCCAGCTCAAGTGTA------AGTAACATGCACTGGTATCAGCAGAAGTCAAGCAC 119
          ||| ||||||||||||||||      ||| || |||||||||| |||||||||||| |  |
Db     72 CAGGGCCAGCTCAAGTGTAAGTTCCAGTTACTTGCACTGGTACCAGCAGAAGTCAGGTGC 131

Qy    120 CTCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTT 179
          ||||||||||||||||||||||| |||||||||  |||||||||||||||| | ||||||
Db    132 CTCCCCCAAACTCTGGGTTTATGGCACATCCAACTTGGCTTCTGGAGTCCCTGCTCGCTT 191

Qy    180 CAGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGA 239
          ||||||| |||||||||| | ||||||||||||||| ||||||||  |||||||||||||
Db    192 CAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGTGTGGAGGCTGAAGA 251

Qy    240 TGCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGGAC 299
          ||||||||||||||| ||  | |||   ||||| |||||| ||||||||||  | |||||
Db    252 TGCTGCCACTTATTACTGCCAGCAGTACAGTGGTTACCCACTCACGTTCGGTGCTGGGAC 311

Qy    300 AAAGTTGGAAATAAAAGGTTCTACCTCCGGATCTGGTAAATCTTCTGAAGGTAAAGGTGT 359
           ||| | ||  |  |||||    | || || |||||      ||| |||  ||||  |
Db    312 CAAGCTTGAGCTGGAAGGTAAATCTTCTGGTTCTGG------TTCCGAATCTAAATCTAC 365

Qy    360 GCAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGTCCTG 419
           |||||| || |||| ||   ||  || ||                 |     |  | ||
Db    366 TCAGCTGAAGGAGTCAGGACCTGTCCTGGTGGCGCCCTCACAGAGCCTGTCCATCACTTG 425

Qy    420 CAAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCCTGA 479
          ||  |  ||||  ||     | |   |||||  | | ||||||||    |||   || |
Db    426 CACCGTCTCTGGGTTTTCATTAACCAACTATGGTGTACACTGGGTTCGCCAGCCTCCAGG 485

Qy    480 ACAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATGACCC 539
          | |||| ||||| ||| | |||   ||     |||    ||| || ||    ||| |  |
Db    486 AAAGGGTCTGGAGTGGCTGGGAGTAATATGGGCTGG---TGGAAACACAAATTATAATTC 542

Qy    540 GAAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTACCTTCA 599
               ||  |  ||     || || |    |||||  |||   | |   |  | | |  |
Db    543 AGCTCTCATGTCCAGACTGAGCATCAGCAAAGACAATTCCAAGAGCCAAGTTTTCTTAAA 602

Qy    600 GCTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGCTTCTTATTACTACTA 659
            | | ||| |||  |  ||| ||||| ||| | || ||||||||           |
Db    603 AATGAACAGTCTGCAAAATGATGACACAGCCATATACTACTGTGCCAAACGACTGGAACG 662

Qy    660 TAGTGCTTACTATGCTATGTACTACTGGGGTCAAGGAACCTCGGTCACCGTCTCCTCA 717
           |    ||||||||||||| |||| ||||||||||||||||| ||||||||||||| |
Db    663 AATCTTTTACTATGCTATGGACTATTGGGGTCAAGGAACCTCAGTCACCGTCTCCTAA 720


RESULT  11
T13739
ID   T13739 standard; DNA; 720 BP.
AC   T13739;
DT   11-OCT-1996  (first entry)
DE   Single chain binding molecule 18-2-3/TRY202' DNA.
KW   Antibody engineering; single polypeptide chain binding molecule;
KW   single chain antibody; SCA; heavy chain; light chain;
KW   monoclonal antibody; MAb; immunoaffinity purification;
KW   18-2-3/TRY202'; ss.
OS   Chimeric Mus sp;
OS   Chimeric synthetic.
FH   Key              Location/Qualifiers
FT   cds              619. .621
FT                    /*tag= a
FT                    /transl_except= (619. .621, aa:Ile)
FT   cds              658. .660
FT                    /*tag= b
FT                    /transl_except= (658. .660, aa:Gly)
PN   US5534621-A.
PD   09-JUL-1996.
PF   02-SEP-1986; 902971.
PR   02-SEP-1986; US-902971.
PR   02-SEP-1987; US-092110.
PR   19-JAN-1989; US-299617.
PR   25-APR-1990; US-512910.
PR   01-APR-1993; US-040440.
PR   06-JUN-1995; US-468992.
PA   (ENZO-) ENZON LABS INC.
PI   Bird RE,  Hardman K,  Ladner RC;
DR   WPI; 96-333309/33.
DR   P-PSDB; R99649.
PT   Immuno:purificn. using single binding chain molecule including
PT   antigen-binding parts of antibody light and heavy chain variable
PT   regions connected by a linker - is smaller, stabler and less
PT   expensive than complete antibodies
PS   Example 9; Fig 39; 78pp; English.
CC   A DNA construct (T13739) codes for single chain binding molecule
CC   18-2-3/TRY202' (R99649), in which VL and VH regions of anti-
CC   fluorescein monoclonal antibody 18-2-3 are joined by a peptide
CC   linker composed primarily of alternating Gly and Ser residues, with
CC   Glu and Lys residues to enhance solubility.  The construct was
CC   expressed in E.coli.  Soluble, folded chain antibody (SCA) was
CC   obtd. which was capable of exhibiting a biological binding
CC   activity equivalent in specificity and affinity to that of a
CC   monoclonal antibody.  The SCA has the advantages of smaller size,
CC   greater stability and reduced cost.
SQ   Sequence   720 BP;    180 A;    184 C;    175 G;    181 T;


  Query Match             44.7%;  Score 322;  DB 1;  Length 720;
  Best Local Similarity   68.7%;  Pred. No. 8e-77;
  Matches  493;  Conservative    0;  Mismatches  210;  Indels   15;  Gaps    3;

Qy      6 GCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATGACCTG 65
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12 GCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATGACCTG 71

Qy     66 CAGTGCCAGCTCAAGTGTA------AGTAACATGCACTGGTATCAGCAGAAGTCAAGCAC 119
          ||| |||||||||||||||      ||| || |||||||||| |||||||||||| |  |
Db     72 CAGGGCCAGCTCAAGTGTAAGTTCCAGTTACTTGCACTGGTACCAGCAGAAGTCAGGTGC 131

Qy    120 CTCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTT 179
          ||||||||||||||||||||||| |||||||||  |||||||||||||||| | ||||||
Db    132 CTCCCCCAAACTCTGGGTTTATGGCACATCCAACTTGGCTTCTGGAGTCCCTGCTCGCTT 191

Qy    180 CAGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGA 239
          ||||||| |||||||||| | ||||||||||||||| ||||||||  |||||||||||||
Db    192 CAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGTGTGGAGGCTGAAGA 251

Qy    240 TGCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGGAC 299
          ||||||||||||||| ||  | |||   ||||| |||||| ||||||||||  | |||||
Db    252 TGCTGCCACTTATTACTGCCAGCAGTACAGTGGTTACCCACTCACGTTCGGTGCTGGGAC 311

Qy    300 AAAGTTGGAAATAAAAGGTTCTACCTCCGGATCTGGTAAATCTTCTGAAGGTAAAGGTGT 359
           ||| | ||  |  |||||    | || || |||||      ||| |||  ||||  |
Db    312 CAAGCTTGAGCTGGAAGGTAAATCTTCTGGTTCTGG------TTCCGAATCTAAATCTAC 365

Qy    360 GCAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGTCCTG 419
           |||||| || |||| ||   ||  || ||                 |     |  | ||
Db    366 TCAGCTGAAGGAGTCAGGACCTGTCCTGGTGGCGCCCTCACAGAGCCTGTCCATCACTTG 425

Qy    420 CAAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCCTGA 479
          ||  |  ||||  ||     | |   |||||  | | ||||||||    |||   || |
Db    426 CACTGTCTCTGGGTTTTCATTAACCAACTATGGTGTACACTGGGTTCGCCAGCCTCCAGG 485

Qy    480 ACAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATGACCC 539
          | |||| ||||| ||| | |||   ||     |||    ||| || ||    ||| |  |
Db    486 AAAGGGTCTGGAGTGGCTGGGAGTAATATGGGCTGG---TGGAAACACAAATTATAATTC 542

Qy    540 GAAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTACCTTCA 599
               ||  |  ||     || || |    |||||  |||   | |   |  | | |  |
Db    543 AGCTCTCATGTCCAGACTGAGCATCAGCAAAGACAATTCCAAGAGCCAAGTTTTCTTAAA 602

Qy    600 GCTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGCTTCTTATTACTACTA 659
            | | ||| |||  |  ||| ||||| ||| | || ||||||||           |
Db    603 AATGAACAGTCTGCAAAATGATGACACAGCCATATACTACTGTGCCAAACGACTGGAACG 662
```

```
Qy    660 TAGTGCTTACTATGCTATGTACTACTGGGGTCAAGGAACCTCGGTCACCGTCTCCTCA 717
           |    ||||||||||||| |||| ||||||||||||||||| ||||||||||||| |
Db    663 AATCTTTTACTATGCTATGGACTATTGGGGTCAAGGAACCTCAGTCACCGTCTCCTAA 720

RESULT  12
T36463
ID   T36463 standard; DNA; 720 BP.
AC   T36463;
DT   13-NOV-1996  (first entry)
DE   18-2-3/TRY202' single chain binding protein-encoding DNA.
KW   Antibody engineering; monoclonal antibody; MAb;
KW   single chain antibody; immunoassay; fluorescein; 18-2-3/TRY202';
KW   single chain binding protein; ss.
OS   Chimeric Mus musculus;
OS   Chimeric synthetic.
FH   Key             Location/Qualifiers
FT   cds             658. .660
FT                   /*tag=  a
FT                   /transl_except= (658. .660, aa:Gly)
PN   US5518889-A.
PD   21-MAY-1996.
PF   02-SEP-1986; 902971.
PR   02-SEP-1986; US-902971.
PR   02-SEP-1987; US-092110.
PR   19-JAN-1989; US-299617.
PR   25-APR-1990; US-512910.
PR   01-APR-1993; US-040440.
PR   06-JUN-1995; US-468988.
PA   (ENZO-) ENZON LABS INC.
PI   Bird RE,  Hardman K,  Ladner RC;
DR   WPI; 96-259060/26.
DR   P-PSDB; W02191.
PT   Immunoassay using single chain antigen binding mol. - as replacement
PT   for labelled or immobilised antibody, are less immunogenic, easier
PT   to engineer, more stable and less expensive
PS   Example 9; Fig 39A-B; 78pp; English.
CC   Computer-designed construct 18-2-3/TRY202' (W02191), encoded by
CC   the DNA sequence given in T36463, comprises the variable regions
CC   of the light and heavy chains of anti-fluorescein monoclonal
CC   antibody (Mab) 18-2-3 linked by a peptide designed to fit into a
CC   groove on the backside of the variable domain structure.  The
CC   18-2-3 VH and VL regions were obtd. by PCR amplification of
CC   hybridoma RNA, and the DNA construct was inserted into vector
CC   pGX3703 and introduced into E. coli.  18-2-3/TRY202' was
CC   expressed as a single chain molecule.  It exhibited biological
CC   binding activity equivalent in specifically and affinity to
CC   that of the original MAb.
SQ   Sequence   720 BP;    180 A;    184 C;    175 G;    181 T;


  Query Match             44.7%;  Score 322;  DB 1;  Length 720;
  Best Local Similarity   68.7%;  Pred. No. 8e-77;
  Matches  493;  Conservative    0;  Mismatches  210;  Indels   15;  Gaps    3;

Qy      6 GCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATGACCTG 65
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12 GCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATGACCTG 71

Qy     66 CAGTGCCAGCTCAAGTGTA------AGTAACATGCACTGGTATCAGCAGAAGTCAAGCAC 119
          ||| |||||||||||||||      ||| || |||||||||| |||||||||||| |  |
Db     72 CAGGGCCAGCTCAAGTGTAAGTTCCAGTTACTTGCACTGGTACCAGCAGAAGTCAGGTGC 131

Qy    120 CTCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTT 179
          ||||||||||||||||||||||| |||||||||  |||||||||||||||| | ||||||
Db    132 CTCCCCCAAACTCTGGGTTTATGGCACATCCAACTTGGCTTCTGGAGTCCCTGCTCGCTT 191

Qy    180 CAGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGA 239
          ||||||| |||||||||| | ||||||||||||||| ||||||||  |||||||||||||
Db    192 CAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGTGTGGAGGCTGAAGA 251

Qy    240 TGCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGGAC 299
          ||||||||||||||| ||  | |||   ||||| |||||| ||||||||||  | |||||
Db    252 TGCTGCCACTTATTACTGCCAGCAGTACAGTGGTTACCCACTCACGTTCGGTGCTGGGAC 311

Qy    300 AAAGTTGGAAATAAAAGGTTCTACCTCCGGATCTGGTAAATCTTCTGAAGGTAAAGGTGT 359
           ||| | ||  |  |||||    | || || |||||      ||| |||  ||||  |
Db    312 CAAGCTTGAGCTGGAAGGTAAATCTTCTGGTTCTGG------TTCCGAATCTAAATCTAC 365

Qy    360 GCAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGTCCTG 419
           |||||| || |||| ||   ||  || ||                 |     |  | ||
Db    366 TCAGCTGAAGGAGTCAGGACCTGTCCTGGTGGCGCCCTCACAGAGCCTGTCCATCACTTG 425

Qy    420 CAAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCCTGA 479
          ||  |  ||||  ||     | |   |||||  | | ||||||||    |||   || |
Db    426 CACTGTCTCTGGGTTTTCATTAACCAACTATGGTGTACACTGGGTTCGCCAGCCTCCAGG 485

Qy    480 ACAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATGACCC 539
          | |||| ||||| ||| | |||   ||     |||    ||| || ||    ||| |  |
Db    486 AAAGGGTCTGGAGTGGCTGGGAGTAATATGGGCTGG---TGGAAACACAAATTATAATTC 542

Qy    540 GAAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTACCTTCA 599
               ||  |  ||     || || |    |||||  |||   | |   |  | | |  |
Db    543 AGCTCTCATGTCCAGACTGAGCATCAGCAAAGACAATTCCAAGAGCCAAGTTTTCTTAAA 602

Qy    600 GCTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGCTTCTTATTACTACTA 659
            | | ||| |||  |  ||| ||||| ||| | || ||||||||           |
Db    603 AATGAACAGTCTGCAAAATGATGACACAGCCATATACTACTGTGCCAAACGACTGGAACG 662

Qy    660 TAGTGCTTACTATGCTATGTACTACTGGGGTCAAGGAACCTCGGTCACCGTCTCCTCA 717
           |    ||||||||||||| |||| ||||||||||||||||| |||||||||||||| |
Db    663 AATCTTTTACTATGCTATGGACTATTGGGGTCAAGGAACCTCAGTCACCGTCTCCTAA 720
```

```
RESULT  13
Q05714
ID   Q05714 standard; DNA; 724 BP.
AC   Q05714;
DT   07-JAN-1991  (first entry)
DE   18-2-3-/TRY202'.
KW   Monoclonal antibody; variable chain; anti-fluorescein; ss.
PN   US4946778-A.
PD   07-AUG-1990.
PF   19-JAN-1989; 299617.
PR   02-SEP-1986; US-902971.
PR   02-SEP-1987; US-092110.
PR   19-JAN-1989; US-299617.
PA   (GENE-) GENEX CORP.
PI   Ladner RC, Bird RE, Hardman K;
DR   WPI; 90-260350/34.
DR   P-PSDB; R06482.
PT   Single polypeptide chain binding molecules - having light chain
PT   variable region of antibody linked by peptide to heavy chain
PT   variable region.
PS   Disclosure; Fig 39; 68pp; English.
CC   The sequence was constructed from the variable regions of an anti-
CC   fluorescein  MAb, 18-3-2, an IgM.   The VL and VH cDNA sequences
CC   were synthesised by priming on RNA isolated from hybridoma cells.
CC   The sequence encodes a single chain binding molecule comprising
CC   the variable regions of heavy and light chains linked by peptides.
CC   The peptide linker was computer-designed to fit inyo a groove on
CC   the backside of the variable domain structure and is composed
CC   primarly of alternating Gly and Ser residues. Glu and Lys residues
CC   are included  to enhance solubility.
CC   See also Q05708-Q05719.
SQ   Sequence   724 BP;    180 A;    186 C;    174 G;    184 T;


  Query Match              44.7%;  Score 321.6;  DB 1;  Length 724;
  Best Local Similarity   68.7%;  Pred. No. 1e-76;
  Matches  492;  Conservative    0;  Mismatches  209;  Indels   15;  Gaps    3;

Qy      6 GCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATGACCTG 65
          ||| |||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db     12 GCTCACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCAACATGACCTG 71
```

```
Qy     66 CAGTGCCAGCTCAAGTGTA------AGTAACATGCACTGGTATCAGCAGAAGTCAAGCAC 119
          ||| |||||||||||||||      ||| || |||||||||| |||||||||||| |  |
Db     72 CAGGGCCAGCTCAAGTGTAAGTTCCAGTTACTTGCACTGGTACCAGCAGAAGTCAGGTGC 131

Qy    120 CTCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTT 179
          ||||||||||||||||||||||| |||||||||  |||||||||||||||| | ||||||
Db    132 CTCCCCCAAACTCTGGGTTTATGGCACATCCAACTTGGCTTCTGGAGTCCCTGCTCGCTT 191

Qy    180 CAGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGA 239
          ||||||| |||||||||| | ||||||||||||||| ||||||||  |||||||||||||
Db    192 CAGTGGCGGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGTGTGGAGGCTGAAGA 251

Qy    240 TGCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGGAC 299
          ||||||||||||||| ||  | |||   ||||| |||||| ||||||||||  | |||||
Db    252 TGCTGCCACTTATTACTGCCAGCAGTACAGTGGTTACCCACTCACGTTCGGTGCTGGGAC 311

Qy    300 AAAGTTGGAAATAAAAGGTTCTACCTCCGGATCTGGTAAATCTTCTGAAGGTAAAGGTGT 359
           ||| | ||  |  |||||    | || || |||||      ||| |||  ||||  |
Db    312 CAAGCTTGAGCTGGAAGGTAAATCTTCTGGTTCTGG------TTCCGAATCTAAATCTAC 365

Qy    360 GCAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGTCCTG 419
           |||||| || |||| ||   ||  || ||                 |     |  | ||
Db    366 TCAGCTGAAGGAGTCAGGACCTGTCCTGGTGGCGCCCTCACAGAGCCTGTCCATCACTTG 425

Qy    420 CAAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCCTGA 479
          ||  |  ||||  ||     | |   |||||  | | ||||||||    |||   || |
Db    426 CACTGTCTCTGGGTTTTCATTAACCAACTATGGTGTACACTGGGTTCGCCAGCCTCCAGG 485

Qy    480 ACAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATGACCC 539
          | |||| ||||| ||| | |||   ||     |||    ||| || ||    ||| |  |
Db    486 AAAGGGTCTGGAGTGGCTGGGAGTAATATGGGCTGG---TGGAAACACAAATTATAATTC 542

Qy    540 GAAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTACCTTCA 599
               ||  |  ||    ||| || |    |||||  |||   | |   |  | | |  |
Db    543 AGCTCTCATGTCCAGACTCAGCATCAGCAAAGACAATTCCAAGAGCCAAGTTTTCTTAAA 602

Qy    600 GCTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGCTTCTTATTACTACTA 659
            | | ||| |||  || ||| ||||| ||| | || ||||||||           |
Db    603 AATGAACAGTCTGCAAATTGATGACACAGCCATATACTACTGTGCCAAACGACTGGAACG 662

Qy    660 TAGTGCTTACTATGCTATGTACTACTGGGGTCAAGGAACCTCGGTCACCGTCTCCT 715
           |    ||||||||||||| |||| ||||||||||||||||| ||||||||||||||
Db    663 AATCTTTTACTATGCTATGGACTATTGGGGTCAAGGAACCTCAGTCACCGTCTCCT 718
```

```
RESULT  14
Q51538
ID   Q51538 standard; DNA; 741 BP.
AC   Q51538;
DT   23-MAY-1994  (first entry)
DE   Coding sequence of polypeptide with affinity for BGH.
KW   Monoclonal antibody; MAb; affinity; binding; antigen; diagnostics;
KW   therapy; imaging; purification; biosensors; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_difference 25. .27
FT                   /*tag=  a
FT                   /transl_except= CGA encodes Proline.
FT   misc_difference 166. .168
FT                   /*tag=  b
FT                   /transl_except= TTC encodes Leucine.
FT   misc_difference 274. .276
FT                   /*tag=  c
FT                   /transl_except= TAC encodes Glutamine.
FT   misc_difference 277. .279
FT                   /*tag=  d
FT                   /transl_except= AGT encodes Tyrosine.
FT   misc_difference 280. .282
FT                   /*tag=  e
FT                   /transl_except= GGT encodes Serine.
FT   misc_difference 484. .486
FT                   /*tag=  f
FT                   /transl_except= CGA Glutamic acid.
```

```
PN   US5260203-A.
PD   09-NOV-1993.
PF   02-SEP-1986; 902971.
PR   02-SEP-1986; US-902971.
PR   02-SEP-1987; US-092110.
PR   19-JAN-1989; US-299617.
PR   25-APR-1990; US-512910.
PA   (ENZO-) ENZON INC.
PI   Bird RE,  Hardman K,  Ladner RC;
DR   WPI; 93-367875/46.
DR   P-PSDB; R43677.
PT   Single chain poly:peptide for binding antigen - comprising light
PT   and heavy chain antigen binding portions linked by peptide linker
PS   Example 3; Figure 28; 78pp; English.
CC   The single chain polypeptide is derived from the mature light and
CC   heavy chains of a mouse monoclonal antibody (MAb) and has affinity
CC   for a given antigen (Bovine growth hormone).  It comprises a first
CC   polypeptide comprising the antigen binding portion of of the light
CC   chain variable region of an antibody and a second polypeptide
CC   comprising the antigen binding portion of the heavy chain variable
CC   region of an antibody and at least one peptide linker linking the
CC   first and second polypeptide chains.  The resulting single chain
CC   polypeptide can be used in diagnostics, therapy
CC   (in vivo and in vitro), imaging, purifications and biosensors.
CC   This particular single chain binding molecule was designated TRY59
CC   and contains one linker peptide.
SQ   Sequence   741 BP;    172 A;    188 C;    192 G;    189 T;


  Query Match             44.6%;  Score 321.2;  DB 1;  Length 741;
  Best Local Similarity   69.0%;  Pred. No. 1.3e-76;
  Matches  504;  Conservative    0;  Mismatches  208;  Indels   18;  Gaps    4;

Qy         6 GCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATGACCTG 65
             ||| |||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db        12 GCTCACCCAGTCTCGAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATGACCTG 71

Qy        66 CAGTGCCAGCTCAAGTGTA------AGTAACATGCACTGGTATCAGCAGAAGTCAAGCAC 119
             ||| |||||||||||||||      ||| || |||||||||  |||||||||||| |  |
Db        72 CAGGGCCAGCTCAAGTGTAAGTTCCAGTTACTTGCACTGGTTCCAGCAGAAGTCAGGTGC 131

Qy       120 CTCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTT 179
             |||||||||||||||| |||||  |||||||||  | |||||||||||||| | ||||||
Db       132 CTCCCCCAAACTCTGGATTTATAGCACATCCAACTTCGCTTCTGGAGTCCCTGCTCGCTT 191

Qy       180 CAGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGA 239
             |||||||||||||||||| | ||||||||||||||| ||||||||  |||||||||||||
Db       192 CAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGTGTGGAGGCTGAAGA 251

Qy       240 TGCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGGAC 299
             ||||||||||||||| ||  |  |  |  |||| |||||| ||||||||||  | |||||
Db       252 TGCTGCCACTTATTACTGCCAGTACAGTGGTGGTTACCCACTCACGTTCGGTGCTGGGAC 311

Qy       300 AAAGTTGGAAATAAAAGGTTC---TACCTCCGGATCTGGTAAATCTTCTGAAGGTAAAGG 356
              ||| || ||  |   |||||   | | || | |            | |     |   |
Db       312 CAAGCTGAAAGAATCTGGTTCTGTTTCTTCTGAACAGCTGGCTCAGTTTCGTTCTCTGGA 371

Qy       357 TGTGCAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGTC 416
             ||||||||||  | |||||||||  ||  | ||       || |  |   | ||  | ||
Db       372 TGTGCAGCTGGTGGAGTCTGGGGGAGACTTAGTGAAGCCTGGAGGGTCCCTGAAACTCTC 431

Qy       417 CTGCAAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCC 476
             |||   ||| ||||  ||||   | |    ||||   ||    |||||    |  |  ||
Db       432 CTGTGCAGCCTCTGGATTCACTTTCATTAGCTATGGCATGTCTTGGGTTCGCCGAACTCC 491

Qy       477 TGAACAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATGA 536
              ||  || | ||||| ||| | | |   |||  |  ||  | |  | | ||    |||
Db       492 AGACAAGAGGCTGGAGTGGGTCGCAACCATTAGTAGTGGTAGTACTTACACCTACTATCC 551

Qy       537 CCCGAAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTACCT 596
                 |   |   ||||         ||   ||   |   |||||   ||   ||| |     |||||
Db       552 AGACAGTGTGAAGGGGCGATTCACCATCTCCAGAGACAATGCCAAGAACACCCTGTACCT 611
```

```
Qy       597 TCAGCTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGC--TTCTTATTAC 654
              ||  | ||| | ||||  ||||||||||| ||| | |||||||||||    |  |||||
Db       612 GCAAATGAGCGGTCTGAAGTCTGAGGACACAGCCATGTATTACTGTGCAAGACGGATTAC 671

Qy       655 TACTATAGT-------GCTTACTATGCTATGTACTACTGGGGTCAAGGAACCTCGGTCAC 707
             |||  ||||       | ||||||||||||| |||||||||||||||||||||| |||||
Db       672 TACGGTAGTACTTACGGATTACTATGCTATGGACTACTGGGGTCAAGGAACCTCAGTCAC 731

Qy       708 CGTCTCCTCA 717
             |||||||| |
Db       732 CGTCTCCTAA 741


RESULT  15
T13737
ID   T13737 standard; DNA; 741 BP.
AC   T13737;
DT   11-OCT-1996  (first entry)
DE   Single chain binding molecule TRY59 DNA.
KW   Antibody engineering; single polypeptide chain binding molecule;
KW   heavy chain; light chain; monoclonal antibody; MAb;
KW   bovine growth hormone; bGH; immunoaffinity purification; TRY59; ss.
OS   Chimeric Mus sp.;
OS   Chimeric synthetic.
FH   Key             Location/Qualifiers
FT   cds             25. .27
FT                   /*tag=  a
FT                   /transl_except= (25. .27, aa:Pro)
FT   cds             166.168
FT                   /*tag=  b
FT                   /transl_except= (166. .168, aa:Leu)
FT   cds             274. .276
FT                   /*tag=  c
FT                   /transl_except= (274. .276, aa:Gln)
FT   cds             277. .279
FT                   /*tag=  d
FT                   /transl_except= (277. .279, aa:Tyr)
FT   cds             280. .282
FT                   /*tag=  e
FT                   /transl_except= (280. .282, aa:Ser)
PN   US5534621-A.
PD   09-JUL-1996.
PF   02-SEP-1986; 902971.
PR   02-SEP-1986; US-902971.
PR   02-SEP-1987; US-092110.
PR   19-JAN-1989; US-299617.
PR   25-APR-1990; US-512910.
PR   01-APR-1993; US-040440.
PR   06-JUN-1995; US-468992.
PA   (ENZO-) ENZON LABS INC.
PI   Bird RE,  Hardman K,  Ladner RC;
DR   WPI; 96-333309/33.
DR   P-PSDB; R99647.
PT   Immuno:purificn. using single binding chain molecule including
PT   antigen-binding parts of antibody light and heavy chain variable
PT   regions connected by a linker - is smaller, stabler and less
PT   expensive than complete antibodies
PS   Example 3; Fig 28; 78pp; English.
CC   A DNA construct (T13737) codes for single chain binding molecule
CC   TRY59 (R99647), in which amino acids (aa) 1-105 of the VL region
CC   of anti-bovine growth hormone mouse IgG1 monoclonal antibody 3C2
CC   and aa2-123 of the 3C2 VH are joined by a peptide linker from the
CC   Fv region of the IgA anti-phosphorylcholine myeloma antibody MCPC-603.
CC   The construct was expressed in E.coli.  Single chain binding
CC   molecules retain the specificity of the heavy and light chain
CC   aggregate variable region of the native antibody but have the
CC   advantages of smaller size, greater stability and reduced cost.
SQ   Sequence   741 BP;    172 A;    188 C;    192 G;    189 T;


  Query Match             44.6%;  Score 321.2;  DB 1;  Length 741;
  Best Local Similarity   69.0%;  Pred. No. 1.3e-76;
  Matches  504;  Conservative    0;  Mismatches  208;  Indels   18;  Gaps    4;
```

```
Qy      6 GCTGACCCAGTCTCCAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATGACCTG 65
          ||| |||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db     12 GCTCACCCAGTCTCGAGCAATCATGTCTGCATCTCCAGGGGAAAAGGTCACCATGACCTG 71

Qy     66 CAGTGCCAGCTCAAGTGTA------AGTAACATGCACTGGTATCAGCAGAAGTCAAGCAC 119
          ||| |||||||||||||||      ||| || |||||||||  |||||||||||| |  |
Db     72 CAGGGCCAGCTCAAGTGTAAGTTCCAGTTACTTGCACTGGTTCCAGCAGAAGTCAGGTGC 131

Qy    120 CTCCCCCAAACTCTGGGTTTATGACACATCCAAACTGGCTTCTGGAGTCCCAGGTCGCTT 179
          |||||||||||||||| |||||  |||||||||  | |||||||||||||| | ||||||
Db    132 CTCCCCCAAACTCTGGATTTATAGCACATCCAACTTCGCTTCTGGAGTCCCTGCTCGCTT 191

Qy    180 CAGTGGCAGTGGGTCTGGAAACTCTTACTCTCTCACGATCAGCAGCATGGAGGCTGAAGA 239
          |||||||||||||||||| | ||||||||||||||| ||||||||  |||||||||||||
Db    192 CAGTGGCAGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGTGTGGAGGCTGAAGA 251

Qy    240 TGCTGCCACTTATTATTGTTATCAGGGGAGTGGGTACCCATTCACGTTCGGCTCGGGGAC 299
          ||||||||||||||| ||  |  |  |  |||| |||||| ||||||||||  | |||||
Db    252 TGCTGCCACTTATTACTGCCAGTACAGTGGTGGTTACCCACTCACGTTCGGTGCTGGGAC 311

Qy    300 AAAGTTGGAAATAAAAGGTTC---TACCTCCGGATCTGGTAAATCTTCTGAAGGTAAAGG 356
           ||| || ||  |   |||||   | | || | |            | |     |   |
Db    312 CAAGCTGAAAGAATCTGGTTCTGTTTCTTCTGAACAGCTGGCTCAGTTTCGTTCTCTGGA 371

Qy    357 TGTGCAGCTGCAGCAGTCTGGGGTTGAGCTTGTCCGAGGAGGGGCCTTAGTCAAGTTGTC 416
          ||||||||||  | |||||||||  ||  | ||       || |  |   | ||  | ||
Db    372 TGTGCAGCTGGTGGAGTCTGGGGGAGACTTAGTGAAGCCTGGAGGGTCCCTGAAACTCTC 431

Qy    417 CTGCAAAGCTTCTGACTTCAACATTAAAGACTATTATATCCACTGGGTGAAGCAGAGGCC 476
          |||   ||| ||||  ||||   | |    ||||   ||    |||||    |  |  ||
Db    432 CTGTGCAGCCTCTGGATTCACTTTCATTAGCTATGGCATGTCTTGGGTTCGCCGAACTCC 491

Qy    477 TGAACAGGGCCTGGAATGGATTGGATGGATTCATCCTGAGAATGGTAATACTGTATATGA 536
           ||  || | ||||| ||| | | |   |||  |  ||  | |  | | ||    |||
Db    492 AGACAAGAGGCTGGAGTGGGTCGCAACCATTAGTAGTGGTAGTACTTACACCTACTATCC 551

Qy    537 CCCGAAATTCCAGGGCAAGGCCAGTATAACAGCAGACACATCCTCCAACGCGGCCTACCT 596
              |   |  ||||      ||  ||  |   |||||   ||   ||| |    |||||
Db    552 AGACAGTGTGAAGGGGCGATTCACCATCTCCAGAGACAATGCCAAGAACACCCTGTACCT 611

Qy    597 TCAGCTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGC--TTCTTATTAC 654
           ||  | ||| | ||||  ||||||||||| ||| | |||||||||||    |  |||||
Db    612 GCAAATGAGCGGTCTGAAGTCTGAGGACACAGCCATGTATTACTGTGCAAGACGGATTAC 671

Qy    655 TACTATAGT-------GCTTACTATGCTATGTACTACTGGGGTCAAGGAACCTCGGTCAC 707
          |||  ||||       | ||||||||||||| |||||||||||||||||||||| |||||
Db    672 TACGGTAGTACTTACGGATTACTATGCTATGGACTACTGGGGTCAAGGAACCTCAGTCAC 731

Qy    708 CGTCTCCTCA 717
          |||||||| |
Db    732 CGTCTCCTAA 741


Search completed: May 13, 1999, 09:55:30
Job time: 1188 sec
```

```
                    GenCore version 4.5
          Copyright (c) 1993 - 1998  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         May 13, 1999, 09:40:48 ; Search time 21.53 Seconds
                                         (without alignments)
                                         412.360 Million cell updates/sec

Title:          US-08-704-178-3
Perfect score:  1267
Sequence:       1 MDLQLTQSPAILSASPGEKV..........RRVTDWYFDVWGAGTTVTVS 237

Scoring table:  BLOSUM62

Searched:       116738 seqs, 37460341 residues

Database :      PIR_58:*
                1:  pir1:*
                2:  pir2:*
                3:  pir3:*
                4:  pir4:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                               SUMMARIES
                 %
Result         Query
  No.   Score  Match Length DB  ID                 Description
-------------------------------------------------------------------------
    1   615.5   48.6    287  3  PC4402             hapten-specific si
    2     540   42.6    122  2  PH0887             Ig heavy chain V r
    3     498   39.3    128  2  I37267             Ig heavy chain V r
    4   496.5   39.2    117  1  MHMS4E             Ig heavy chain V r
    5     495   39.1    128  2  C37267             Ig heavy chain V r
    6     490   38.7    107  2  PC4405             anti-AFP antibody
    7     490   38.7    120  2  E45722             anti-glycoprotein
    8   487.5   38.5    117  1  MHMSJ5             Ig heavy chain V r
    9     482   38.0    106  2  PL0082             Ig kappa chain V r
   10   480.5   37.9    137  2  H32513             Ig heavy chain pre
   11     477   37.6    120  2  F45722             anti-glycoprotein
   12     475   37.5    114  2  S26319             Ig heavy chain V r
   13   474.5   37.5    135  2  PS0057             Ig heavy chain pre
   14   473.5   37.4    119  2  PH0099             Ig heavy chain V r
   15     473   37.3    119  2  F30502             Ig heavy chain V r
   16     471   37.2    118  1  MHMS38             Ig heavy chain V r
   17   470.5   37.1    125  2  PH0100             Ig heavy chain V r
   18     468   36.9    107  2  PD0011             immunoglobulin VLB
   19     466   36.8    107  2  A42848             Ig light chain V r
   20   461.5   36.4    121  2  F37266             Ig heavy chain V r
   21     461   36.4    128  2  A37267             Ig heavy chain V r
   22     460   36.3    107  2  A30562             Ig kappa chain V r
   23     460   36.3    105  2  S26338             Ig kappa chain V r
   24   457.5   36.1    108  2  G30560             Ig kappa chain V r
   25   457.5   36.1    118  2  PL0200             anti-DNA autoantib
   26   455.5   36.0    108  2  S38720             Ig light chain V r
   27     455   35.9    107  2  B30562             Ig kappa chain V r
   28     455   35.9    122  2  E37267             Ig heavy chain V r
   29     454   35.8    139  2  A27609             Ig heavy chain pre
   30   452.5   35.7    119  2  B53285             Ig heavy chain V a
   31     452   35.7    140  2  PL0013             Ig kappa chain pre
   32   447.5   35.3    235  2  S25058             Ig kappa chain - m
   33     447   35.3     97  2  S26341             Ig light chain V r
   34   445.5   35.2    121  2  H37266             Ig heavy chain V r
   35     445   35.1    106  2  PS0071             Ig kappa chain V r
   36     445   35.1    103  2  S29591             Ig kappa chain V r
   37     445   35.1    112  2  S09957             Ig heavy chain V-D
   38     445   35.1    113  2  S55534             Ig heavy chain V r
   39     444   35.0    130  1  JL0079             Ig kappa chain pre
```

```
   40     443   35.0    130  2  B32456             Ig kappa chain pre
   41   442.5   34.9    123  2  B30560             Ig heavy chain V r
   42   441.5   34.8    117  2  S03305             Ig heavy chain V r
   43     441   34.8    104  2  B49049             Ig light chain V r
   44     441   34.8    104  2  JC6076             anti-D-dimer monoc
   45     441   34.8    113  2  S55528             Ig heavy chain V r

                              ALIGNMENTS


RESULT   1
PC4402
hapten-specific single-chain antibody variable fragment and alkaline phosphatase fusi
C;Species: Escherichia coli
C;Date: 03-Sep-1998 #sequence_revision 03-Sep-1998 #text_change 03-Sep-1998
C;Accession: PC4402
R;Suzuki, C.; Ueda, H.; Suzuki, E.; Nagamune, T.
J. Biochem. 122, 322-329, 1997
A;Title: Construction, bacterial expression, and characterization of hapten-specific
A;Reference number: PC4402
A;Accession: PC4402
A;Status: preliminary
A;Residues: 1-287 <SUZ>


  Query Match             48.6%;  Score 615.5;  DB 3;  Length 287;
  Best Local Similarity   51.0%;  Pred. No. 3.5e-38;
  Matches  132;  Conservative   32;  Mismatches   70;  Indels   25;  Gaps    6;

Qy       1 MDLQ--LTQSPAILSASPGEKVTMTCR----ATPSVSYMHWYQQKPGSSPKPWIYTTSNL 54
           ||:|  :||  | |: |||| ||:|||    |  : :| :| |:||       |  |:|
Db      23 MDIQAVVTQESA-LTTSPGETVTLTCRSSTGAVTTSNYANWVQEKPDHLFTGLIGGTNNR 81

Qy      55 ASGVPARFSGGGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIKGSTSGSG 114
           | ||||||||   |   :||::  : || | |:|  |  :   ||||:|| :  |   :
Db      82 APGVPARFSGSLIGDKAALTITGAQTEDEAIYFCALWYSNHWVFGGGTKLTVLSSADDAK 141

Qy     115 KSSEGKG------------VQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHG 162
           |    |             ||||: | |:|||| |:|:||| |||:|| : |:||||  |
Db     142 KDDAKKDDAKKDDAKKDGQVQLQQPGAELVKPGASVKLSCKASGYTFTSYWMHWVKQRPG 201

Qy     163 KNLEWIGLINPYNGDTNYNQKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARRVTD 222
           : ||||| |:| :| | ||:||| ||| |||| ||||||:| ||||||||||||||   |
Db     202 RGLEWIGRIDPNSGGTKYNEKFKSKATLTVDKPSSTAYMQLSSLTSEDSAVYYCAR--YD 259

Qy     223 W----YFDVWGAGTTVTVS 237
           :    ||| || |||:|||
Db     260 YYGSSYFDYWGQGTTLTVS 278


RESULT   2
PH0887
Ig heavy chain V region (anti-CD3) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 09-Oct-1992 #sequence_revision 09-Oct-1992 #text_change 16-Aug-1996
C;Accession: PH0887
R;Shalaby, M.R.; Shepard, H.M.; Presta, L.; Rodrigues, M.L.; Beverley, P.C.L.; Feldma
J. Exp. Med. 175, 217-225, 1992
A;Title: Development of humanized bispecific antibodies reactive with cytotoxic lymph
A;Reference number: PH0885; MUID:92113462
A;Accession: PH0887
A;Molecule type: mRNA
A;Residues: 1-122 <SHA>
A;Note: the authors translated the codon TTC for residue 70 as Leu
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin


  Query Match             42.6%;  Score 540;  DB 2;  Length 122;
  Best Local Similarity   85.0%;  Pred. No. 4.2e-33;
  Matches  102;  Conservative    6;  Mismatches    8;  Indels    4;  Gaps    1;
```

```
Qy    122 VQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGDTNYN 181
          ||||:||||:|||| ||||||| |||||||:|||||||||||||||:|||||| | : ||
Db      2 VQLQQSGPELVKPGASMKISCKASGYSFTGYTMNWVKQSHGKNLEWMGLINPYKGVSTYN 61

Qy    182 QKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARR----VTDWYFDVWGAGTTVTVS 237
          |||| ||||||||||||||||||||||||||||||||      :||||||||||||||||
Db     62 QKFKDKATFTVDKSSSTAYMELLSLTSEDSAVYYCARSGYYGDSDWYFDVWGAGTTVTVS 121


RESULT   3
I37267
Ig heavy chain V region (129) - mouse
C;Species: Mus musculus (house mouse)
C;Date: 18-Oct-1991 #sequence_revision 18-Oct-1991 #text_change 16-Aug-1996
C;Accession: I37267
R;Ruff-Jamison, S.; Campos-Gonzalez, R.; Glenney Jr., J.R.
J. Biol. Chem. 266, 6607-6613, 1991
A;Title: Heavy and light chain variable region sequences and antibody properties of anti
A;Reference number: A38740; MUID:91177923
A;Accession: I37267
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-128 <RUF>
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin
F;19-102/Domain: immunoglobulin homology <IMM>


  Query Match             39.3%;  Score 498;  DB 2;  Length 128;
  Best Local Similarity   77.2%;  Pred. No. 5.2e-30;
  Matches   95;  Conservative   10;  Mismatches   12;  Indels    6;  Gaps    1;

Qy    121 GVQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGDTNY 180
          |||||:||||:|||| |:|||||||||:|| :||:||||||||:||||| ||| || |:|
Db      5 GVQLQQSGPELVKPGASVKISCKTSGYTFTEYTMHWVKQSHGKSLEWIGGINPNNGGTSY 64

Qy    181 NQKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARR------VTDWYFDVWGAGTTV 234
          ||||||||| ||||||||||||| |||||||||||||||         :||| || |||:
Db     65 NQKFKGKATLTVDKSSSTAYMELRSLTSEDSAVYYCARRGLTTVVAKSYYFDYWGQGTTL 124

Qy    235 TVS 237
          |||
Db    125 TVS 127


RESULT   4
MHMS4E
Ig heavy chain V region (MOPC 104E) - mouse
C;Species: Mus musculus (house mouse)
C;Date: 30-Nov-1979 #sequence_revision 14-Nov-1983 #text_change 16-Aug-1996
C;Accession: A02039
R;Kehry, M.R.; Fuhrman, J.S.; Schilling, J.W.; Rogers, J.; Sibley, C.H.; Hood, L.E.
Biochemistry 21, 5415-5424, 1982
A;Title: Complete amino acid sequence of a mouse mu chain: homology among heavy chain co
A;Reference number: A02039; MUID:83075344
A;Accession: A02039
A;Molecule type: protein
A;Residues: 1-117 <KEH>
C;Comment: The glycosylated asparagine residue does not have the usual N-X-S/T context a
C;Comment: This protein binds dextran.
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: glycoprotein; heterotetramer; immunoglobulin
F;15-98/Domain: immunoglobulin homology <IMM>
F;22-96/Disulfide bonds: #status predicted
F;55/Binding site: carbohydrate (Asn) (covalent) #status atypical


  Query Match             39.2%;  Score 496.5;  DB 1;  Length 117;
  Best Local Similarity   81.0%;  Pred. No. 6e-30;
  Matches   94;  Conservative    9;  Mismatches   12;  Indels    1;  Gaps    1;

Qy    122 VQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGDTNYN 181
          ||||:||||:|||| |:|:||| |||:|| : | ||||||||:||||| ||| || |:||
```

```
Db      2 VQLQQSGPELVKPGASVKMSCKASGYTFTDYYMKWVKQSHGKSLEWIGDINPNNGGTSYN 61

Qy    182 QKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARRVTDWYFDVWGAGTTVTVS 237
          |||||||| |||||||||||:| ||||||||||||||   ||||||||||||||||
Db     62 QKFKGKATLTVDKSSSTAYMQLNSLTSEDSAVYYCARDY-DWYFDVWGAGTTVTVS 116

RESULT   5
C37267
Ig heavy chain V region (Py69) - mouse
C;Species: Mus musculus (house mouse)
C;Date: 18-Oct-1991 #sequence_revision 18-Oct-1991 #text_change 16-Aug-1996
C;Accession: C37267
R;Ruff-Jamison, S.; Campos-Gonzalez, R.; Glenney Jr., J.R.
J. Biol. Chem. 266, 6607-6613, 1991
A;Title: Heavy and light chain variable region sequences and antibody properties of a
A;Reference number: A38740; MUID:91177923
A;Accession: C37267
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-128 <RUF>
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin


  Query Match             39.1%;  Score 495;  DB 2;  Length 128;
  Best Local Similarity   76.4%;  Pred. No. 8.5e-30;
  Matches   94;  Conservative   10;  Mismatches   13;  Indels    6;  Gaps    1;

Qy    121 GVQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGDTNY 180
          |||||:||||:|||| |||||||||||:|| :||:||||||||:|||||  || :| : |
Db      5 GVQLQQSGPELVKPGASMKISCKTSGYTFTEYTMHWVKQSHGKSLEWIGRFNPNSGGSTY 64

Qy    181 NQKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARR------VTDWYFDVWGAGTTV 234
          ||||||||| ||||||||||||| |||||||||||||||       | :||| || |||:
Db     65 NQKFKGKATLTVDKSSSTAYMELRSLTSEDSAVYYCARRGPYGNYYTSYYFDYWGQGTTL 124

Qy    235 TVS 237
          |||
Db    125 TVS 127


RESULT   6
PC4405
anti-AFP antibody F3 light chain variable region - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 10-Nov-1997 #sequence_revision 23-Jan-1998 #text_change 26-Feb-1998
C;Accession: PC4405
R;Deng, J.B.; Han, H.; Su, C.Z.; Chen, C.Q.
Chinese Biochem. J. 12, 648-653, 1996
A;Title: Generation of a phage display library of the immunoglobulin repertoire from
A;Reference number: PC4405
A;Accession: PC4405
A;Molecule type: mRNA
A;Residues: 1-107 <DEN>
A;Experimental source: spleen cell
C;Superfamily: immunoglobulin V region; immunoglobulin homology


  Query Match             38.7%;  Score 490;  DB 2;  Length 107;
  Best Local Similarity   84.0%;  Pred. No. 1.6e-29;
  Matches   89;  Conservative    9;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      2 DLQLTQSPAILSASPGEKVTMTCRATPSVSYMHWYQQKPGSSPKPWIYTTSNLASGVPAR 61
          |::|||||||:|||||||||||| |: |:|||||||||||:||| ||| || ||||||||
Db      1 DIELTQSPAIMSASPGEKVTMTCSASSSISYMHWYQQKPGTSPKRWIYDTSKLASGVPAR 60

Qy     62 FSGGGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIK 107
          ||| |||||||||:||||||||||||||||  :|||||||:|||||
Db     61 FSGSGSGTSYSLTISRVEAEDAATYYCQQWRDNPPTFGGGTKLEIK 106


RESULT   7
```

```
E45722
anti-glycoprotein H monoclonal antibody heavy-chain variable domain (Mab 5) - mouse (fra
C;Species: Mus musculus (house mouse)
C;Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-May-1997
C;Accession: E45722
R;Simpson, J.A.; Chow, J.C.; Baker, J.; Avdalovic, N.; Yuan, S.; Au, D.; Co, M.S.; Vasqu
J. Virol. 67, 489-496, 1993
A;Title: Neutralizing monoclonal antibodies that distinguish three antigenic sites on hu
A;Reference number: A45722; MUID:93100833
A;Accession: E45722
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-120 <SIM>
A;Note: sequence extracted from NCBI backbone (NCBIP:120593)
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: glycoprotein


  Query Match             38.7%;  Score 490;  DB 2;  Length 120;
  Best Local Similarity   80.5%;  Pred. No. 1.8e-29;
  Matches   95;  Conservative    8;  Mismatches   13;  Indels    2;  Gaps    2;

Qy     122 VQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGL-INPYNGDTNY 180
           ||||:||||:|||| ||||||| | |||||:||||||||||:||||||| |||||| |:|
Db       2 VQLQQSGPELVKPGASMKISCKASVYSFTGYTMNWVKQSHGQNLEWIGLIINPYNGGTSY 61

Qy     181 NQKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARR-VTDWYFDVWGAGTTVTVS 237
           ||||||||| ||||||:||||||||||| ||||||| ||   |:  | || ||:||||
Db      62 NQKFKGKATLTVDKSSNTAYMELLSLTSADSAVYYCTRRGFRDYSMDYWGQGTSVTVS 119


RESULT   8
MHMSJ5
Ig heavy chain V region (J558) - mouse
C;Species: Mus musculus (house mouse)
C;Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Aug-1996
C;Accession: A26242
R;Schilling, J.; Clevinger, B.; Davie, J.M.; Hood, L.
Nature 283, 35-40, 1980
A;Title: Amino acid sequence of homogeneous antibodies to dextran and DNA rearrangements
A;Reference number: A26242; MUID:80078170
A;Accession: A26242
A;Molecule type: protein
A;Residues: 1-117 <SCH>
A;Note: the sequences of 10 hybridoma proteins that also bind dextran differ from that s
C;Comment: This protein binds dextran.
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; hybridoma; immunoglobulin
F;15-98/Domain: immunoglobulin homology <IMM>
F;22-96/Disulfide bonds: #status predicted


  Query Match             38.5%;  Score 487.5;  DB 1;  Length 117;
  Best Local Similarity   80.2%;  Pred. No. 2.7e-29;
  Matches   93;  Conservative    9;  Mismatches   13;  Indels    1;  Gaps    1;

Qy     122 VQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGDTNYN 181
           ||||:||||:|||| |:|:||| |||:|| : | ||||||||:||||| ||| || |:||
Db       2 VQLQQSGPELVKPGASVKMSCKASGYTFTDYYMKWVKQSHGKSLEWIGDINPNNGGTSYN 61

Qy     182 QKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARRVTDWYFDVWGAGTTVTVS 237
           |||||||| |||||||||||:| ||||||||||||||    |||||||||||||||
Db      62 QKFKGKATLTVDKSSSTAYMQLNSLTSEDSAVYYCARD-RYWYFDVWGAGTTVTVS 116


RESULT   9
PL0082
Ig kappa chain V region (2D3) - mouse
C;Species: Mus musculus (house mouse)
C;Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 16-Aug-1996
C;Accession: PL0082
R;Meek, K.; Hasemann, C.; Pollok, B.; Alkan, S.S.; Brait, M.; Slaoui, M.; Urbain, J.; Ca
J. Exp. Med. 169, 519-533, 1989
A;Title: Structural characterization of antiidiotypic antibodies; evidence that Ab2s
A;Reference number: PL0080; MUID:89094248
A;Accession: PL0082
A;Molecule type: mRNA
A;Residues: 1-106 <MEE>
A;Experimental source: strain BALB/c
A;Note: the sequence shown here is from the V kappa region of an antiidiotypic monocl
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin


  Query Match             38.0%;  Score 482;  DB 2;  Length 106;
  Best Local Similarity   87.4%;  Pred. No. 6.1e-29;
  Matches   90;  Conservative    6;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       5 LTQSPAILSASPGEKVTMTCRATPSVSYMHWYQQKPGSSPKPWIYTTSNLASGVPARFSG 64
           |:||||||||||||||||||||: |||||:|||||||||||||   ||||||||||||||
Db       4 LSQSPAILSASPGEKVTMTCRASSSVSYMNWYQQKPGSSPKPWESATSNLASGVPARFSG 63

Qy      65 GGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIK 107
            |||||||||:||||||||||||| ||| :|||||||:|||||
Db      64 SGSGTSYSLTISRVEAEDAATYYCNQWSSNPPTFGGGTKLEIK 106


RESULT  10
H32513
Ig heavy chain precursor V region (BXW16) - mouse
C;Species: Mus musculus (house mouse)
C;Date: 21-May-1990 #sequence_revision 31-Dec-1990 #text_change 20-Mar-1998
C;Accession: H32513
R;Kofler, R.; Strohal, R.; Balderas, R.S.; Johnson, M.E.; Noonan, D.J.; Duchosal, M.A
J. Clin. Invest. 82, 852-860, 1988
A;Title: Immunoglobulin kappa light chain variable region gene complex organization a
A;Reference number: A94689; MUID:88331394
A;Accession: H32513
A;Molecule type: DNA
A;Residues: 1-137 <KOF>
A;Cross-references: GB:M20831; NID:g196949; PID:g196950
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin


  Query Match             37.9%;  Score 480.5;  DB 2;  Length 137;
  Best Local Similarity   74.4%;  Pred. No. 1e-28;
  Matches   93;  Conservative   11;  Mismatches   18;  Indels    3;  Gaps    1;

Qy     113 SGKSSEGKGVQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLIN 172
           :|  ||   :|||:|| |:|||| |:||||| |||||||: ||||||||||:||||| ||
Db      15 AGVHSE---IQLQQSGAELVKPGASVKISCKASGYSFTGYNMNWVKQSHGKSLEWIGNIN 71

Qy     173 PYNGDTNYNQKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARRVTDWYFDVWGAGT 232
           || | |:|||||||||| |||||||||||:| ||||||||||||||:     || || ||
Db      72 PYYGSTSYNQKFKGKATLTVDKSSSTAYMQLNSLTSEDSAVYYCARKNYGSSFDYWGQGT 131

Qy     233 TVTVS 237
           |:|||
Db     132 TLTVS 136


RESULT  11
F45722
anti-glycoprotein H monoclonal antibody heavy-chain variable domain (Mab 33) - mouse
C;Species: Mus musculus (house mouse)
C;Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-May-1997
C;Accession: F45722
R;Simpson, J.A.; Chow, J.C.; Baker, J.; Avdalovic, N.; Yuan, S.; Au, D.; Co, M.S.; Va
J. Virol. 67, 489-496, 1993
A;Title: Neutralizing monoclonal antibodies that distinguish three antigenic sites on
A;Reference number: A45722; MUID:93100833
A;Accession: F45722
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-120 <SIM>
```

```
A;Note: sequence extracted from NCBI backbone (NCBIP:120594)
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: glycoprotein


  Query Match             37.6%;  Score 477;  DB 2;  Length 120;
  Best Local Similarity   77.1%;  Pred. No. 1.6e-28;
  Matches   91;  Conservative   11;  Mismatches   14;  Indels    2;  Gaps    2;

Qy    122 VQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGL-INPYNGDTNY 180
          ||||:||||:|||| ||||||| |||||||:||||||| |||::||||| ||| || |:|
Db      2 VQLQQSGPELVKPGASMKISCKASGYSFTGYTMNWVKQRHGKSIEWIGLIINPNNGGTSY 61

Qy    181 NQKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARR-VTDWYFDVWGAGTTVTVS 237
          ||:|:|| | |:|||||| ||||||||||||||||||||   |:  | || ||:||||
Db     62 NQRFRGKTTLTLDKSSSTVYMELLSLTSEDSAVYYCARRGFRDYALDSWGQGTSVTVS 119


RESULT  12
S26319
Ig heavy chain V region - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 19-Mar-1998 #sequence_revision 19-Mar-1998 #text_change 17-Apr-1998
C;Accession: S26319
R;Stark, S.E.; Caton, A.J.
J. Exp. Med. 174, 613-624, 1991
A;Title: Antibodies that are specific for a single amino acid interchange in a protein e
A;Reference number: S26309
A;Accession: S26319
A;Molecule type: mRNA
A;Residues: 1-114 <STA>
A;Cross-references: EMBL:X59172
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin


  Query Match             37.5%;  Score 475;  DB 2;  Length 114;
  Best Local Similarity   78.4%;  Pred. No. 2.1e-28;
  Matches   91;  Conservative    9;  Mismatches   10;  Indels    6;  Gaps    2;

Qy    125 QESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGDTNYNQKF 184
          |:||||:|||| |:||||| |||||||: ||||||||||:||||| |||||||| |||||
Db      1 QQSGPELVKPGASVKISCKASGYSFTGYFMNWVKQSHGKSLEWIGRINPYNGDTFYNQKF 60

Qy    185 KGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARRVTDWYF---DVWGAGTTVTVS 237
          ||||| |||||||||:|||||||||||||||| :   | |:   | || ||:||||
Db     61 KGKATLTVDKSSSTAHMELLSLTSEDSAVYYCGK---DGYYGAMDYWGQGTSVTVS 113


RESULT  13
PS0057
Ig heavy chain precursor V region (PAR) - mouse
C;Species: Mus musculus (house mouse)
C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 08-Sep-1997
C;Accession: PS0057
R;Yaoita, Y.; Takahashi, M.; Azuma, C.; Kanai, Y.; Honjo, T.
J. Biochem. 104, 337-343, 1988
A;Title: Biased expression of variable region gene families of the immunoglobulin heavy
A;Reference number: PS0057; MUID:89197817
A;Accession: PS0057
A;Molecule type: DNA
A;Residues: 1-135 <YAO>
A;Cross-references: GB:D00307; NID:g220448; PID:d1000661; PID:g220449
A;Note: the authors translated the codon AAG for residue 32 as Asn and GAC for 92 as Gly
C;Comment: The gene encoding this protein was isolated from a hybridoma that produces an
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-135/Product: Ig heavy chain V region PAR #status predicted <MAT>


  Query Match             37.5%;  Score 474.5;  DB 2;  Length 135;
  Best Local Similarity   76.0%;  Pred. No. 2.8e-28;
  Matches   95;  Conservative    9;  Mismatches   16;  Indels    5;  Gaps    2;

Qy    113 SGKSSEGKGVQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLIN 172
          :|  ||   ||||:||||:|||| |:||||| |||||||: ||||||||||:||||| ||
Db     15 AGVFSE---VQLQQSGPELVKPGASVKISCKASGYSFTGYFMNWVKQSHGKSLEWIGRIN 71

Qy    173 PYNGDTNYNQKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARRVTDWYFDVWGAGT 232
          |||||| |||||||||| |||||||||:||| ||||||||||||||   ::    || ||
Db     72 PYNGDTFYNQKFKGKATLTVDKSSSTAHMELRSLTSEDSAVYYCAR--LNYRGAYWGQGT 129

Qy    233 TVTVS 237
           ||||
Db    130 LVTVS 134


RESULT  14
PH0099
Ig heavy chain V region (anti-cyclosporin F) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 16-Aug-1996
C;Accession: PH0099
R;Schmitter, D.; Poch, O.; Zeder, G.; Heinrich, G.F.; Kocher, H.P.; Quesniaux, V.F.J.
Mol. Immunol. 27, 1029-1038, 1990
A;Title: Analysis of the structural diversity of monoclonal antibodies to cyclosporin
A;Reference number: PH0087; MUID:91042649
A;Accession: PH0099
A;Molecule type: mRNA
A;Residues: 1-119 <SCH>
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin
F;31-35/Region: complementarity-determining 1
F;50-66/Region: complementarity-determining 2
F;99-106/Region: complementarity-determining 3


  Query Match             37.4%;  Score 473.5;  DB 2;  Length 119;
  Best Local Similarity   78.0%;  Pred. No. 2.9e-28;
  Matches   92;  Conservative    9;  Mismatches   12;  Indels    5;  Gaps    2;

Qy    122 VQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGDTNYN 181
          ||||:||||:|:|| ||||||| ||||| |:|| |||||||||||||||||||:|     |
Db      2 VQLQQSGPELVEPGASMKISCKASGYSFAGYTMIWVKQSHGKNLEWIGLINPYDGSIATN 61

Qy    182 QKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARRVTDWYF--DVWGAGTTVTVS 237
          |||||||| |||||||||||:||||||||||||||||    :|:  | || ||:||||
Db     62 QKFKGKATLTVDKSSSTAYMDLLSLTSEDSAVYYCAR---GYYYAVDYWGQGTSVTVS 116


RESULT  15
F30502
Ig heavy chain V region (A52) - mouse
C;Species: Mus musculus (house mouse)
C;Date: 03-Nov-1988 #sequence_revision 03-Nov-1988 #text_change 16-Aug-1996
C;Accession: F30502
R;Eilat, D.; Webster, D.M.; Rees, A.R.
J. Immunol. 141, 1745-1753, 1988
A;Title: V region sequences of anti-DNA and anti-RNA autoantibodies from NZB/NZW F-1
A;Reference number: A30502; MUID:88315787
A;Accession: F30502
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-119 <EIL>
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin


  Query Match             37.3%;  Score 473;  DB 2;  Length 119;
  Best Local Similarity   77.1%;  Pred. No. 3.1e-28;
  Matches   91;  Conservative   11;  Mismatches   14;  Indels    2;  Gaps    2;

Qy    122 VQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGDTNYN 181
          :|||:|| |:|||| |:||||| |||||||: ||||||||||:||||| |||| | |:||
Db      2 IQLQQSGAELVKPGASVKISCKASGYSFTGYNMNWVKQSHGKSLEWIGKINPYYGSTSYN 61
```

```
Qy    182 QKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCAR-RV-TDWYFDVWGAGTTVTVS 237
          |||||:|| ||||||||||:| |||||||||||||| |:    ||| || |||:|||
Db     62 QKFKGQATLTVDKSSSTAYMQLNSLTSEDSAVYYCARGRLRRGGYFDYWGQGTTLTVS 119


Search completed: May 13, 1999, 10:03:48
Job time: 1380 sec
```

```
                         GenCore version 4.5
                Copyright (c) 1993 - 1998  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         May 13, 1999, 09:55:33 ; Search time 16.53 Seconds
                                          (without alignments)
                                          384.825 Million cell updates/sec

Title:          US-08-704-178-3
Perfect score:  1267
Sequence:       1 MDLQLTQSPAILSASPGEKV..........RRVTDWYFDVWGAGTTVTVS 237

Scoring table:  BLOSUM62

Searched:       74019 seqs, 26840295 residues

Database :      SwissProt_36:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                                SUMMARIES
                 %
Result           Query
  No.   Score    Match Length DB  ID                   Description
--------------------------------------------------------------------------
    1   496.5    39.2    117  1  HV12_MOUSE           P01756 mus musculu
    2   487.5    38.5    117  1  HV13_MOUSE           P01757 mus musculu
    3     471    37.2    118  1  HV51_MOUSE           P06330 mus musculu
    4   439.5    34.7    136  1  HV15_MOUSE           P01759 mus musculu
    5     437    34.5    107  1  KV6F_MOUSE           P04940 mus musculu
    6     432    34.1    107  1  KV6I_MOUSE           P04943 mus musculu
    7     431    34.0    107  1  KV6G_MOUSE           P04941 mus musculu
    8     431    34.0    107  1  KV6H_MOUSE           P04942 mus musculu
    9     430    33.9    107  1  KV6J_MOUSE           P04944 mus musculu
   10     429    33.9    108  1  KV6K_MOUSE           P04945 mus musculu
   11   424.5    33.5    120  1  HV03_MOUSE           P01747 mus musculu
   12     422    33.3    117  1  HV14_MOUSE           P01758 mus musculu
   13     420    33.1    107  1  KV6D_MOUSE           P01678 mus musculu
   14     419    33.1    107  1  KV6A_MOUSE           P01675 mus musculu
   15     414    32.7    107  1  KV6B_MOUSE           P01676 mus musculu
   16   412.5    32.6    139  1  HV07_MOUSE           P01751 mus musculu
   17     412    32.5    107  1  KV6C_MOUSE           P01677 mus musculu
   18     407    32.1    140  1  HV02_MOUSE           P01746 mus musculu
   19     407    32.1    120  1  HV50_MOUSE           P06329 mus musculu
   20   403.5    31.8    137  1  HV11_MOUSE           P01755 mus musculu
   21     401    31.6    129  1  KV4A_MOUSE           P01680 mus musculu
   22     397    31.3    107  1  KV6E_MOUSE           P01679 mus musculu
   23   381.5    30.1    138  1  HV48_MOUSE           P03980 mus musculu
   24     381    30.1    117  1  HV06_MOUSE           P01750 mus musculu
   25   373.5    29.5    121  1  HV01_MOUSE           P01745 mus musculu
   26     373    29.4    117  1  HV04_MOUSE           P01748 mus musculu
   27     371    29.3    117  1  HV52_MOUSE           P06327 mus musculu
   28   363.5    28.7    117  1  HV09_MOUSE           P01753 mus musculu
   29     362    28.6    117  1  HV05_MOUSE           P01749 mus musculu
   30   360.5    28.5    111  1  KV3T_MOUSE           P01672 mus musculu
   31     360    28.4    117  1  HV49_MOUSE           P06328 mus musculu
   32   355.5    28.1    111  1  KV3H_MOUSE           P01660 mus musculu
   33   355.5    28.1    111  1  KV3R_MOUSE           P01670 mus musculu
   34     355    28.0    117  1  HV10_MOUSE           P01754 mus musculu
   35   353.5    27.9    111  1  KV3J_MOUSE           P01662 mus musculu
   36   353.5    27.9    111  1  KV3L_MOUSE           P01664 mus musculu
   37   352.5    27.8    131  1  KV3I_MOUSE           P01661 mus musculu
   38   351.5    27.7    111  1  KV3O_MOUSE           P01667 mus musculu
   39   347.5    27.4    108  1  KV1B_HUMAN           P01594 homo sapien
   40   346.5    27.3    111  1  KV3M_MOUSE           P01665 mus musculu
   41   346.5    27.3    111  1  KV3S_MOUSE           P01671 mus musculu
   42   345.5    27.3    111  1  KV3K_MOUSE           P01663 mus musculu
   43   345.5    27.3    111  1  KV3N_MOUSE           P01666 mus musculu
```

```
   44     344    27.2    109  1  KV3B_HUMAN           P01620 homo sapien
   45   343.5    27.1    111  1  KV3Q_MOUSE           P01669 mus musculu

                                ALIGNMENTS

RESULT   1
HV12_MOUSE
ID   HV12_MOUSE     STANDARD;      PRT;   117 AA.
AC   P01756;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1991 (REL. 20, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN V REGION (MOPC 104E).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 83075344.
RA   KEHRY M.R., FUHRMAN J.S., SCHILLING J.W., ROGERS J., SIBLEY C.H.,
RA   HOOD L.E.;
RL   BIOCHEMISTRY 21:5415-5424(1982).
CC   -!- THE SEQUENCE OF THE LIGHT CHAIN OF THIS IGM MYELOMA PROTEIN HAS
CC       ALSO BEEN DETERMINED.
CC   -!- THIS PROTEIN BINDS DEXTRAN.
DR   PIR; A02039; MHMS4E.
DR   HSSP; P01772; 1FGV.
KW   IMMUNOGLOBULIN V REGION; GLYCOPROTEIN.
FT   DISULFID     22     96       BY SIMILARITY.
FT   CARBOHYD     55     55
FT   NON_TER     117    117
SQ   SEQUENCE   117 AA;  12983 MW;  1F6CC304 CRC32;


  Query Match             39.2%;  Score 496.5;  DB 1;  Length 117;
  Best Local Similarity   81.0%;  Pred. No. 5.6e-31;
  Matches   94;  Conservative    9;  Mismatches   12;  Indels    1;  Gaps    1;

Qy    122 VQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGDTNYN 181
          ||||:||||:|||| |:|:||| |||:|| : | ||||||||:||||| ||| || |:||
Db      2 VQLQQSGPELVKPGASVKMSCKASGYTFTDYYMKWVKQSHGKSLEWIGDINPNNGGTSYN 61

Qy    182 QKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARRVTDWYFDVWGAGTTVTVS 237
          |||||||| |||||||||||:| ||||||||||||||   ||||||||||||||||
Db     62 QKFKGKATLTVDKSSSTAYMQLNSLTSEDSAVYYCARDY-DWYFDVWGAGTTVTVS 116


RESULT   2
HV13_MOUSE
ID   HV13_MOUSE     STANDARD;      PRT;   117 AA.
AC   P01757;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1991 (REL. 20, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN V REGION (J558).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 80078170.
RA   SCHILLING J., CLEVINGER B., DAVIE J.M., HOOD L.;
RL   NATURE 283:35-40(1980).
CC   -!- THE SEQUENCES OF 10 HYBRIDOMA PROTEINS THAT ALSO BIND DEXTRAN
CC       DIFFER FROM THAT SHOWN AT 1-7 POSITIONS, MANY OF WHICH OCCUR IN
CC       THE D AND J SEGMENTS.
CC   -!- THIS PROTEIN BINDS DEXTRAN.
DR   PIR; A26242; MHMSJ5.
DR   HSSP; P01789; 2FGW.
KW   IMMUNOGLOBULIN V REGION.
FT   DISULFID     22     96       BY SIMILARITY.
FT   NON_TER      117    117
SQ   SEQUENCE   117 AA;  13024 MW;  E7548A05 CRC32;

  Query Match             38.5%;  Score 487.5;  DB 1;  Length 117;
  Best Local Similarity   80.2%;  Pred. No. 2.7e-30;
  Matches   93;  Conservative    9;  Mismatches   13;  Indels    1;  Gaps    1;

Qy     122 VQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGDTNYN 181
           ||||:||||:|||| |:|:||| |||:|| : | ||||||||:||||| ||| || |:||
Db       2 VQLQQSGPELVKPGASVKMSCKASGYTFTDYYMKWVKQSHGKSLEWIGDINPNNGGTSYN 61

Qy     182 QKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARRVTDWYFDVWGAGTTVTVS 237
           |||||||| |||||||||||:| ||||||||||||||    |||||||||||||||
Db      62 QKFKGKATLTVDKSSSTAYMQLNSLTSEDSAVYYCARD-RYWYFDVWGAGTTVTVS 116


RESULT   3
HV51_MOUSE
ID   HV51_MOUSE     STANDARD;      PRT;   118 AA.
AC   P06330;
DT   01-JAN-1988 (REL. 06, CREATED)
DT   01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT   01-MAR-1992 (REL. 21, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN V REGION (AC38 205.12).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 84182519.
RA   DILDROP R., BOVENS J., SIEKEVITZ M., BEYREUTHER K., RAJEWSKY K.;
RL   EMBO J. 3:517-523(1984).
DR   PIR; A02040; MHMS38.
DR   HSSP; P01772; 1FGV.
KW   IMMUNOGLOBULIN V REGION.
FT   DOMAIN         1     98       V SEGMENT.
FT   DOMAIN        99    104       D SEGMENT.
FT   DOMAIN       105    118       J SEGMENT.
FT   DISULFID      22     96       BY SIMILARITY.
FT   NON_TER      118    118
SQ   SEQUENCE   118 AA;  12934 MW;  2D1DCE77 CRC32;

  Query Match             37.2%;  Score 471;  DB 1;  Length 118;
  Best Local Similarity   78.4%;  Pred. No. 4.6e-29;
  Matches   91;  Conservative    7;  Mismatches   18;  Indels    0;  Gaps    0;

Qy     122 VQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGDTNYN 181
           ||||:||||:|||| |:||||| |||:|| : ||||||||||:||||| ||| || |:||
Db       2 VQLQQSGPELVKPGASVKISCKASGYTFTDYYMNWVKQSHGKSLEWIGDINPNNGGTSYN 61

Qy     182 QKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARRVTDWYFDVWGAGTTVTVS 237
           |||||||| |||||||  |||| ||||||||||||||      ||||| |||||||
Db      62 QKFKGKATLTVDKSSSATYMELRSLTSEDSAVYYCARGYGYDPFDVWGTGTTVTVS 117


RESULT   4
HV15_MOUSE
ID   HV15_MOUSE     STANDARD;      PRT;   136 AA.
AC   P01759;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN PRECURSOR V REGION (BCL1).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 82222262.
RA   KNAPP M.R., LIU C.-P., NEWELL N., WARD R.B., TUCKER P.W., STROBER S.,
RA   BLATTNER F.R.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 79:2996-3000(1982).
DR   EMBL; J00494; G195011; -.
DR   PIR; A02042; HVMSB1.
DR   HSSP; P01789; 1JEL.
KW   IMMUNOGLOBULIN V REGION; SIGNAL.
FT   SIGNAL         1     19
FT   CHAIN         20    136       IG HEAVY CHAIN V REGION (BCL1).
FT   NON_TER      136    136
SQ   SEQUENCE   136 AA;  15078 MW;  E04F1C7F CRC32;

  Query Match             34.7%;  Score 439.5;  DB 1;  Length 136;
  Best Local Similarity   72.4%;  Pred. No. 1.2e-26;
  Matches   84;  Conservative   14;  Mismatches   17;  Indels    1;  Gaps    1;

Qy     122 VQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGDTNYN 181
           ||||:||||||:|| |:||||| |||:|| : |:|||||| |:|||||:|: |||:|:||
Db      21 VQLQQSGPEVVRPGVSVKISCKGSGYTFTDYAMHWVKQSHAKSLEWIGVISTYNGNTSYN 80

Qy     182 QKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARRVTDWYFDVWGAGTTVTVS 237
           |||||||| |||||||| :|||  |||||||  ||||   :  ||| || |||:|||
Db      81 QKFKGKATMTVDKSSSTVHMELARLTSEDSANLYCARYYGN-YFDYWGQGTTLTVS 135


RESULT   5
KV6F_MOUSE
ID   KV6F_MOUSE     STANDARD;      PRT;   107 AA.
AC   P04940;
DT   13-AUG-1987 (REL. 05, CREATED)
DT   13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT   01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
DE   IG KAPPA CHAIN V-VI REGION (NQ2-17.4.1).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 83271467.
RA   KAARTINEN M., GRIFFITHS G.M., MARKHAM A.F., MILSTEIN C.;
RL   NATURE 304:320-324(1983).
CC   -!- FUNCTION: ANTI-2-PHENYL OXAZOLONE (PHOX) ANTIBODY.
DR   EMBL; K00735; G196435; -.
DR   HSSP; P01679; 1FVB.
KW   IMMUNOGLOBULIN V REGION; HYBRIDOMA.
FT   DOMAIN         1     23       FRAMEWORK 1.
FT   DOMAIN        24     33       COMPLEMENTARITY-DETERMINING 1.
FT   DOMAIN        34     48       FRAMEWORK 2.
FT   DOMAIN        49     55       COMPLEMENTARITY-DETERMINING 2.
FT   DOMAIN        56     87       FRAMEWORK 3.
FT   DOMAIN        88     96       COMPLEMENTARITY-DETERMINING 3.
FT   DOMAIN        97    106       FRAMEWORK 4.
FT   DISULFID      23     87       BY SIMILARITY.
FT   NON_TER      107    107
SQ   SEQUENCE   107 AA;  11561 MW;  1BC0C1E3 CRC32;

  Query Match             34.5%;  Score 437;  DB 1;  Length 107;
  Best Local Similarity   77.7%;  Pred. No. 1.4e-26;
  Matches   80;  Conservative   11;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       5 LTQSPAILSASPGEKVTMTCRATPSVSYMHWYQQKPGSSPKPWIYTTSNLASGVPARFSG 64
           |||||||:|||||:|||||| |: ||||||||||| |:||| ||| || |||||||||||
Db       4 LTQSPAIMSASPGQKVTMTCSASSSVSYMHWYQQKSGTSPKRWIYDTSKLASGVPARFSG 63

Qy      65 GGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIK 107
            || ||||||:: ::||||||||||||| :| ||| |:|||:|
Db      64 SGSATSYSLTITSMQAEDAATYYCQQWSSNPLTFGAGTKLELK 106


RESULT   6
KV6I_MOUSE
ID   KV6I_MOUSE     STANDARD;      PRT;   107 AA.
AC   P04943;
```

```
DT   13-AUG-1987 (REL. 05, CREATED)
DT   13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT   01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
DE   IG KAPPA CHAIN V-VI REGION (NQ6-8.3.1).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 83271467.
RA   KAARTINEN M., GRIFFITHS G.M., MARKHAM A.F., MILSTEIN C.;
RL   NATURE 304:320-324(1983).
CC   -!- FUNCTION: ANTI-2-PHENYL OXAZOLONE (PHOX) ANTIBODY.
DR   EMBL; K00740; G196445; -.
DR   HSSP; P01679; 1FVB.
KW   IMMUNOGLOBULIN V REGION; HYBRIDOMA.
FT   DOMAIN        1     23       FRAMEWORK 1.
FT   DOMAIN       24     33       COMPLEMENTARITY-DETERMINING 1.
FT   DOMAIN       34     48       FRAMEWORK 2.
FT   DOMAIN       49     55       COMPLEMENTARITY-DETERMINING 2.
FT   DOMAIN       56     87       FRAMEWORK 3.
FT   DOMAIN       88     96       COMPLEMENTARITY-DETERMINING 3.
FT   DOMAIN       97    106       FRAMEWORK 4.
FT   DISULFID     23     87       BY SIMILARITY.
FT   NON_TER     107    107
SQ   SEQUENCE   107 AA;  11572 MW;  2D976197 CRC32;

  Query Match             34.1%;  Score 432;  DB 1;  Length 107;
  Best Local Similarity   76.7%;  Pred. No. 3.3e-26;
  Matches   79;  Conservative   11;  Mismatches   13;  Indels    0;  Gaps    0;

Qy       5 LTQSPAILSASPGEKVTMTCRATPSVSYMHWYQQKPGSSPKPWIYTTSNLASGVPARFSG 64
           |||||||:|||||:|||||| |: ||||||||||| |:||| ||| || |||| ||||||
Db       4 LTQSPAIMSASPGQKVTMTCSASSSVSYMHWYQQKSGTSPKRWIYDTSKLASGXPARFSG 63

Qy      65 GGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIK 107
            || ||||||:: ::||||||||||||| :| ||| |:|||:|
Db      64 SGSATSYSLTITSMQAEDAATYYCQQWSSNPLTFGAGTKLELK 106


RESULT   7
KV6G_MOUSE
ID   KV6G_MOUSE     STANDARD;      PRT;   107 AA.
AC   P04941;
DT   13-AUG-1987 (REL. 05, CREATED)
DT   13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT   01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
DE   IG KAPPA CHAIN V-VI REGION (NQ2-48.2.2).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 83271467.
RA   KAARTINEN M., GRIFFITHS G.M., MARKHAM A.F., MILSTEIN C.;
RL   NATURE 304:320-324(1983).
CC   -!- FUNCTION: ANTI-2-PHENYL OXAZOLONE (PHOX) ANTIBODY.
DR   EMBL; K00737; G196439; -.
DR   HSSP; P01679; 1FVB.
KW   IMMUNOGLOBULIN V REGION; HYBRIDOMA.
FT   DOMAIN        1     23       FRAMEWORK 1.
FT   DOMAIN       24     33       COMPLEMENTARITY-DETERMINING 1.
FT   DOMAIN       34     48       FRAMEWORK 2.
FT   DOMAIN       49     55       COMPLEMENTARITY-DETERMINING 2.
FT   DOMAIN       56     87       FRAMEWORK 3.
FT   DOMAIN       88     96       COMPLEMENTARITY-DETERMINING 3.
FT   DOMAIN       97    106       FRAMEWORK 4.
FT   DISULFID     23     87       BY SIMILARITY.
FT   NON_TER     107    107
SQ   SEQUENCE   107 AA;  11556 MW;  1AB8295D CRC32;

  Query Match             34.0%;  Score 431;  DB 1;  Length 107;
  Best Local Similarity   76.7%;  Pred. No. 4e-26;
  Matches   79;  Conservative   11;  Mismatches   13;  Indels    0;  Gaps    0;

Qy       5 LTQSPAILSASPGEKVTMTCRATPSVSYMHWYQQKPGSSPKPWIYTTSNLASGVPARFSG 64
           |||||||:|||||:|||||| |: ||||||||||| |:||| ||| || |||||||||||
Db       4 LTQSPAIMSASPGQKVTMTCSASSSVSYMHWYQQKSGTSPKRWIYDTSKLASGVPARFSG 63

Qy      65 GGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIK 107
            || ||||||:: ::||||||||||||| :| ||| |:|| :|
Db      64 SGSATSYSLTITSMQAEDAATYYCQQWSSNPLTFGAGTKLXLK 106


RESULT   8
KV6H_MOUSE
ID   KV6H_MOUSE     STANDARD;      PRT;   107 AA.
AC   P04942;
DT   13-AUG-1987 (REL. 05, CREATED)
DT   13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT   01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
DE   IG KAPPA CHAIN V-VI REGION (NQ5-61.1.2).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 83271467.
RA   KAARTINEN M., GRIFFITHS G.M., MARKHAM A.F., MILSTEIN C.;
RL   NATURE 304:320-324(1983).
CC   -!- FUNCTION: ANTI-2-PHENYL OXAZOLONE (PHOX) ANTIBODY.
DR   EMBL; K00739; G196443; -.
DR   HSSP; P01679; 1FVB.
KW   IMMUNOGLOBULIN V REGION; HYBRIDOMA.
FT   DOMAIN        1     23       FRAMEWORK 1.
FT   DOMAIN       24     33       COMPLEMENTARITY-DETERMINING 1.
FT   DOMAIN       34     48       FRAMEWORK 2.
FT   DOMAIN       49     55       COMPLEMENTARITY-DETERMINING 2.
FT   DOMAIN       56     87       FRAMEWORK 3.
FT   DOMAIN       88     96       COMPLEMENTARITY-DETERMINING 3.
FT   DOMAIN       97    106       FRAMEWORK 4.
FT   DISULFID     23     87       BY SIMILARITY.
FT   NON_TER     107    107
SQ   SEQUENCE   107 AA;  11605 MW;  1CF3CE8D CRC32;

  Query Match             34.0%;  Score 431;  DB 1;  Length 107;
  Best Local Similarity   76.7%;  Pred. No. 4e-26;
  Matches   79;  Conservative   11;  Mismatches   13;  Indels    0;  Gaps    0;

Qy       5 LTQSPAILSASPGEKVTMTCRATPSVSYMHWYQQKPGSSPKPWIYTTSNLASGVPARFSG 64
           |||||||:|||||:|||||| |: ||||||||||| |:||| ||| || | |||||||||
Db       4 LTQSPAIMSASPGQKVTMTCSASSSVSYMHWYQQKSGTSPKRWIYDTSKLDSGVPARFSG 63

Qy      65 GGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIK 107
            || ||||||:: ::||||||||||||| :| ||| |:|||:|
Db      64 SGSATSYSLTITSMQAEDAATYYCQQWSSNPLTFGAGTKLELK 106


RESULT   9
KV6J_MOUSE
ID   KV6J_MOUSE     STANDARD;      PRT;   107 AA.
AC   P04944;
DT   13-AUG-1987 (REL. 05, CREATED)
DT   13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT   01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
DE   IG KAPPA CHAIN V-VI REGION (NQ5-78.2.6).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 83271467.
RA   KAARTINEN M., GRIFFITHS G.M., MARKHAM A.F., MILSTEIN C.;
```

```
RL   NATURE 304:320-324(1983).
CC   -!- FUNCTION: ANTI-2-PHENYL OXAZOLONE (PHOX) ANTIBODY.
DR   EMBL; K00744; G196453; -.
DR   HSSP; P01679; 1FVB.
KW   IMMUNOGLOBULIN V REGION; HYBRIDOMA.
FT   DOMAIN        1     23       FRAMEWORK 1.
FT   DOMAIN       24     33       COMPLEMENTARITY-DETERMINING 1.
FT   DOMAIN       34     48       FRAMEWORK 2.
FT   DOMAIN       49     55       COMPLEMENTARITY-DETERMINING 2.
FT   DOMAIN       56     87       FRAMEWORK 3.
FT   DOMAIN       88     96       COMPLEMENTARITY-DETERMINING 3.
FT   DOMAIN       97    106       FRAMEWORK 4.
FT   DISULFID     23     87       BY SIMILARITY.
FT   NON_TER     107    107
SQ   SEQUENCE   107 AA;  11611 MW;  E926BBEB CRC32;

  Query Match             33.9%;  Score 430;  DB 1;  Length 107;
  Best Local Similarity   76.7%;  Pred. No. 4.7e-26;
  Matches   79;  Conservative   10;  Mismatches   14;  Indels    0;  Gaps    0;

Qy      5 LTQSPAILSASPGEKVTMTCRATPSVSYMHWYQQKPGSSPKPWIYTTSNLASGVPARFSG 64
          |||||||:|||||:|||||| |: ||||||||||| |:||| ||| || ||||||||| |
Db      4 LTQSPAIMSASPGQKVTMTCSASSSVSYMHWYQQKSGTSPKRWIYDTSKLASGVPARFXG 63

Qy     65 GGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIK 107
           || ||||||:: ::||||||||||||| :| ||| |:||| |
Db     64 SGSATSYSLTITSMQAEDAATYYCQQWSSNPLTFGSGTKLEXK 106


RESULT  10
KV6K_MOUSE
ID   KV6K_MOUSE     STANDARD;      PRT;   108 AA.
AC   P04945;
DT   13-AUG-1987 (REL. 05, CREATED)
DT   13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT   01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
DE   IG KAPPA CHAIN V-VI REGION (NQ2-6.1).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 83271467.
RA   KAARTINEN M., GRIFFITHS G.M., MARKHAM A.F., MILSTEIN C.;
RL   NATURE 304:320-324(1983).
DR   EMBL; K00746; G554059; -.
DR   HSSP; P01679; 1BAF.
KW   IMMUNOGLOBULIN V REGION; HYBRIDOMA.
FT   DOMAIN        1     23       FRAMEWORK 1.
FT   DOMAIN       24     33       COMPLEMENTARITY-DETERMINING 1.
FT   DOMAIN       34     48       FRAMEWORK 2.
FT   DOMAIN       49     55       COMPLEMENTARITY-DETERMINING 2.
FT   DOMAIN       56     87       FRAMEWORK 3.
FT   DOMAIN       88     98       COMPLEMENTARITY-DETERMINING 3.
FT   DOMAIN       99    108       FRAMEWORK 4.
FT   DISULFID     23     87       BY SIMILARITY.
FT   NON_TER     108    108
SQ   SEQUENCE   108 AA;  11713 MW;  8B48A204 CRC32;

  Query Match             33.9%;  Score 429;  DB 1;  Length 108;
  Best Local Similarity   77.1%;  Pred. No. 5.7e-26;
  Matches   81;  Conservative   11;  Mismatches   11;  Indels    2;  Gaps    1;

Qy      5 LTQSPAILSASPGEKVTMTCRATPSVSYMHWYQQKPGSSPKPWIYTTSNLASGVPARFSG 64
          |||||||:|||||:|||||| |: |||||:||||||||||:  || ||||||||| ||||
Db      4 LTQSPAIMSASPGQKVTMTCSASSSVSYMYWYQQKPGSSPRLLIYDTSNLASGVPVRFSG 63

Qy     65 GGSGTSYSLTVSRVEAEDAATYYCQQWSRSPP--TFGGGSKLEIK 107
           || ||||||::|::|||||||||||||  ||  ||| |:|||:|
Db     64 SGSATSYSLTITRMQAEDAATYYCQQWSSYPPMLTFGAGTKLELK 108
```

```
RESULT  11
HV03_MOUSE
ID   HV03_MOUSE     STANDARD;      PRT;   120 AA.
AC   P01747;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   21-JUL-1986 (REL. 01, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN V REGION (36-65).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 83131846.
RA   SIEKEVITZ M., GEFTER M.L., BRODEUR P., RIBLET R.,
RA   MARSHAK-ROTHSTEIN A.;
RL   EUR. J. IMMUNOL. 12:1023-1032(1982).
CC   -!- FROM ANALYSIS OF THE SIZES OF SEVERAL OTHER DIFFERENTIATED GENES
CC       THAT HYBRIDIZE TO THIS ONE, THE AUTHORS CONCLUDE THAT ALL OF
CC       THESE V REGIONS HAVE REARRANGED TO THE SAME J SEGMENT, JH2.
DR   PIR; A02028; HVMSG7.
DR   HSSP; P01789; 6FAB.
KW   IMMUNOGLOBULIN V REGION; ANTIARSONATE ANTIBODY; HYBRIDOMA.
FT   NON_TER     120    120
SQ   SEQUENCE   120 AA;  13307 MW;  BBA8CCA1 CRC32;

  Query Match             33.5%;  Score 424.5;  DB 1;  Length 120;
  Best Local Similarity   68.1%;  Pred. No. 1.4e-25;
  Matches   81;  Conservative   14;  Mismatches   21;  Indels    3;  Gaps    1;

Qy    122 VQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGDTNYN 181
          ||||:|| |:|: | |:|:||| |||:|| : :|||||  |: ||||| ||| || | ||
Db      1 VQLQQSGAELVRAGSSVKMSCKASGYTFTSYGINWVKQRPGQGLEWIGYINPGNGYTKYN 60

Qy    182 QKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARRV---TDWYFDVWGAGTTVTVS 237
          :||||| | |||||||||||:| ||||||||||:||| |     :||| || |||:|||
Db     61 EKFKGKTTLTVDKSSSTAYMQLRSLTSEDSAVYFCARSVYYGGSYYFDYWGQGTTLTVS 119


RESULT  12
HV14_MOUSE
ID   HV14_MOUSE     STANDARD;      PRT;   117 AA.
AC   P01758;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN PRECURSOR V REGION (108A).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 81245215.
RA   GIVOL D., ZAKUT R., EFFRON K., RECHAVI G., RAM D., COHEN J.B.;
RL   NATURE 292:426-430(1981).
DR   EMBL; J00488; G554028; -.
DR   PIR; A02041; HVMS8A.
DR   HSSP; P01789; 1IGI.
KW   IMMUNOGLOBULIN V REGION; SIGNAL.
FT   SIGNAL        1     19
FT   CHAIN        20    117       IG HEAVY CHAIN V REGION (10A).
FT   NON_TER     117    117
SQ   SEQUENCE   117 AA;  12972 MW;  941C002E CRC32;

  Query Match             33.3%;  Score 422;  DB 1;  Length 117;
  Best Local Similarity   75.5%;  Pred. No. 2.1e-25;
  Matches   80;  Conservative    8;  Mismatches   18;  Indels    0;  Gaps    0;

Qy    113 SGKSSEGKGVQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLIN 172
          || :     ||||:||||:|||| |:||||| |||:|| : |:||||||||:||||| |
Db       12 SGTAGVHSEVQLQQSGPELVKPGASVKISCKASGYTFTDYNMHWVKQSHGKSLEWIGYIY 71

Qy      173 PYNGDTNYNQKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCAR 218
            |||| | |||||| ||| ||| ||||||||| ||||||||||||||
Db       72 PYNGGTGYNQKFKSKATLTVDNSSSTAYMELSSLTSEDSAVYYCAR 117


RESULT  13
KV6D_MOUSE
ID   KV6D_MOUSE     STANDARD;      PRT;   107 AA.
AC   P01678;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
DE   IG KAPPA CHAIN V-VI REGION (SAPC 10).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 81054757.
RA   RUDIKOFF S., RAO D.N., GLAUDEMANS C.P.J., POTTER M.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 77:4270-4274(1980).
CC   -!- THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT BIND GALACTAN.
DR   PIR; A01941; KVMSX4.
DR   HSSP; P01679; 2FBJ.
KW   IMMUNOGLOBULIN V REGION.
FT   DOMAIN        1     23       FRAMEWORK 1.
FT   DOMAIN       24     33       COMPLEMENTARITY-DETERMINING 1.
FT   DOMAIN       34     48       FRAMEWORK 2.
FT   DOMAIN       49     55       COMPLEMENTARITY-DETERMINING 2.
FT   DOMAIN       56     87       FRAMEWORK 3.
FT   DOMAIN       88     96       COMPLEMENTARITY-DETERMINING 3.
FT   DOMAIN       97    106       FRAMEWORK 4.
FT   DISULFID     23     87       BY SIMILARITY.
FT   NON_TER     107    107
SQ   SEQUENCE   107 AA;  11554 MW;  609EDE13 CRC32;


  Query Match              33.1%;  Score 420;  DB 1;  Length 107;
  Best Local Similarity   74.5%;  Pred. No. 2.6e-25;
  Matches   79;  Conservative   11;  Mismatches   16;  Indels    0;  Gaps    0;

Qy        2 DLQLTQSPAILSASPGEKVTMTCRATPSVSYMHWYQQKPGSSPKPWIYTTSNLASGVPAR 61
            :: ||||||| :|| |:|||:|| |: ||||||||||| |:|||||||  | ||||||||
Db        1 EIVLTQSPAITAASLGQKVTITCSASSSVSYMHWYQQKSGTSPKPWIYEISKLASGVPAR 60

Qy       62 FSGGGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIK 107
            ||| ||||||||:| :|||||| ||||||:    |||||:|||||
Db       61 FSGSGSGTSYSLTISSMEAEDAAIYYCQQWNYPLITFGGGTKLEIK 106


RESULT  14
KV6A_MOUSE
ID   KV6A_MOUSE     STANDARD;      PRT;   107 AA.
AC   P01675;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
DE   IG KAPPA CHAIN V-VI REGION (XRPC 44).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 79082830.
RA   RAO D.N., RUDIKOFF S., POTTER M.;
RL   BIOCHEMISTRY 17:5555-5559(1978).
CC   -!- THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT BIND GALACTAN.
DR   PIR; A01941; KVMSX4.
DR   HSSP; P01679; 2FBJ.
KW   IMMUNOGLOBULIN V REGION.
FT   DOMAIN        1     23       FRAMEWORK 1.
```

```
FT   DOMAIN       24     33       COMPLEMENTARITY-DETERMINING 1.
FT   DOMAIN       34     48       FRAMEWORK 2.
FT   DOMAIN       49     55       COMPLEMENTARITY-DETERMINING 2.
FT   DOMAIN       56     87       FRAMEWORK 3.
FT   DOMAIN       88     96       COMPLEMENTARITY-DETERMINING 3.
FT   DOMAIN       97    106       FRAMEWORK 4.
FT   DISULFID     23     87       BY SIMILARITY.
FT   NON_TER     107    107
SQ   SEQUENCE   107 AA;  11627 MW;  2E7F7479 CRC32;

  Query Match              33.1%;  Score 419;  DB 1;  Length 107;
  Best Local Similarity   74.5%;  Pred. No. 3.1e-25;
  Matches   79;  Conservative   11;  Mismatches   16;  Indels    0;  Gaps    0;

Qy        2 DLQLTQSPAILSASPGEKVTMTCRATPSVSYMHWYQQKPGSSPKPWIYTTSNLASGVPAR 61
            :: ||||||| :|| |:|||:|| |: ||||||||||| |:|||||||  | ||||||||
Db        1 EIVLTQSPAITAASLGQKVTITCSASSSVSYMHWYQQKSGTSPKPWIYEISKLASGVPAR 60

Qy       62 FSGGGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIK 107
            ||| ||||||||:| :|||||| ||||||:    |||||:|||||
Db       61 FSGSGSGTSYSLTISSMEAEDAAIYYCQQWNYPLWTFGGGTKLEIK 106


RESULT  15
KV6B_MOUSE
ID   KV6B_MOUSE     STANDARD;      PRT;   107 AA.
AC   P01676;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
DE   IG KAPPA CHAIN V-VI REGION (XRPC 24).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 79082830.
RA   RAO D.N., RUDIKOFF S., POTTER M.;
RL   BIOCHEMISTRY 17:5555-5559(1978).
CC   -!- THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT BIND GALACTAN.
DR   PIR; A01941; KVMSX4.
DR   HSSP; P01679; 2FBJ.
KW   IMMUNOGLOBULIN V REGION.
FT   DOMAIN        1     23       FRAMEWORK 1.
FT   DOMAIN       24     33       COMPLEMENTARITY-DETERMINING 1.
FT   DOMAIN       34     48       FRAMEWORK 2.
FT   DOMAIN       49     55       COMPLEMENTARITY-DETERMINING 2.
FT   DOMAIN       56     87       FRAMEWORK 3.
FT   DOMAIN       88     96       COMPLEMENTARITY-DETERMINING 3.
FT   DOMAIN       97    106       FRAMEWORK 4.
FT   DISULFID     23     87       BY SIMILARITY.
FT   NON_TER     107    107
SQ   SEQUENCE   107 AA;  11584 MW;  91B2CD6E CRC32;


  Query Match              32.7%;  Score 414;  DB 1;  Length 107;
  Best Local Similarity   73.6%;  Pred. No. 7.4e-25;
  Matches   78;  Conservative   11;  Mismatches   17;  Indels    0;  Gaps    0;

Qy        2 DLQLTQSPAILSASPGEKVTMTCRATPSVSYMHWYQQKPGSSPKPWIYTTSNLASGVPAR 61
            :: ||||||| :|| |:|||:|| |: ||||||||||| |:|||||||  | ||||||||
Db        1 EIVLTQSPAITAASLGQKVTITCSASSSVSYMHWYQQKSGTSPKPWIYEISKLASGVPAR 60

Qy       62 FSGGGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIK 107
            ||| ||||||||:| :|||||| ||||||:    ||| |:|||||
Db       61 FSGSGSGTSYSLTISSMEAEDAAIYYCQQWNYPLITFGSGTKLEIK 106


Search completed: May 13, 1999, 10:05:48
Job time: 615 sec
```

```
                        GenCore version 4.5
                Copyright (c) 1993 - 1998  Compugen Ltd.


OM protein - protein search, using sw model

Run on:         May 13, 1999, 09:54:22 ; Search time 28.48 Seconds
                                         (without alignments)
                                         459.098 Million cell updates/sec

Title:          US-08-704-178-3
Perfect score:  1267
Sequence:       1 MDLQLTQSPAILSASPGEKV..........RRVTDWYFDVWGAGTTVTVS 237

Scoring table:  BLOSUM62

Searched:       180763 seqs, 55169189 residues

Database :      SPTREMBL_8:*
                1:  sp_fungi:*
                2:  sp_human:*
                3:  sp_invertebrate:*
                4:  sp_mammal:*
                5:  sp_mhc:*
                6:  sp_organelle:*
                7:  sp_phage:*
                8:  sp_plant:*
                9:  sp_bacteria:*
                10:  sp_rodent:*
                11:  sp_virus:*
                12:  sp_vertebrate:*
                13:  sp_unclassified:*
                14:  sp_archea:*

    Pred. No. is the number of results predicted by chance to have a
    score greater than or equal to the score of the result being printed,
    and is derived by analysis of the total score distribution.

                                SUMMARIES
                    %
Result              Query
  No.     Score     Match Length DB  ID                   Description
--------------------------------------------------------------------------
    1       436      34.4    130 10  P80913                P80913 mus musculu
    2       239      18.9    100  2  Q15535               Q15535 homo sapien
    3     193.5      15.3     88  2  O75737               O75737 homo sapien
    4     182.5      14.4     79  2  O75731               O75731 homo sapien
    5       181      14.3     78  2  O75739               O75739 homo sapien
    6     178.5      14.1     78  2  O75723               O75723 homo sapien
    7     176.5      13.9     82  2  O75729               O75729 homo sapien
    8     175.5      13.9     97  2  O43234               O43234 homo sapien
    9       174      13.7     77  2  O75726               O75726 homo sapien
   10     173.5      13.7     85  2  O75724               O75724 homo sapien
   11       170      13.4     77  2  O75728               O75728 homo sapien
   12     169.5      13.4     81  2  O75721               O75721 homo sapien
   13     168.5      13.3     76  2  O75742               O75742 homo sapien
   14       168      13.3     80  2  O75727               O75727 homo sapien
   15     165.5      13.1     77  2  O75741               O75741 homo sapien
   16       165      13.0     86  2  O75740               O75740 homo sapien
   17     164.5      13.0     82  2  O75725               O75725 homo sapien
   18       164      12.9     72  2  O75738               O75738 homo sapien
   19     163.5      12.9     86  2  O75722               O75722 homo sapien
   20       163      12.9     81  2  O75719               O75719 homo sapien
   21     162.5      12.8     78  2  O75730               O75730 homo sapien
   22     162.5      12.8     75  2  O75743               O75743 homo sapien
   23     161.5      12.7    418 10  O70426                O70426 rattus norv
   24     161.5      12.7    509 10  P97710                P97710 rattus norv
   25       160      12.6     82  2  O75732               O75732 homo sapien
   26     159.5      12.6     81  2  O75734               O75734 homo sapien
   27       152      12.0    398  2  O00241               O00241 homo sapien
   28     150.5      11.9     80  2  O75735               O75735 homo sapien
   29     150.5      11.9     74  2  O75744               O75744 homo sapien
```

```
   30     149.5      11.8     50  2  Q15533               Q15533 homo sapien
   31       144      11.4     76  2  O75733               O75733 homo sapien
   32     143.5      11.3    513 10  P97797                P97797 mus musculu
   33     142.5      11.2    503  2  P78324               P78324 homo sapien
   34     141.5      11.2    509 10  O35924                O35924 mus musculu
   35       139      11.0     45  2  O76051               O76051 homo sapien
   36     137.5      10.9    509 10  O08907                O08907 mus musculu
   37     130.5      10.3    509 10  O88555                O88555 mus musculu
   38       130      10.3    254 12  Q90557               Q90557 ginglymosto
   39     128.5      10.1    700  2  O75054               O75054 homo sapien
   40     128.5      10.1     78  2  O75720               O75720 homo sapien
   41     128.5      10.1    258 12  Q90556                Q90556 ginglymosto
   42     127.5      10.1    145  2  Q16237               Q16237 homo sapien
   43     127.5      10.1    250 12  Q90569                Q90569 ginglymosto
   44       127      10.0     81  2  O75736               O75736 homo sapien
   45       122       9.6    258 12  Q90559                Q90559 ginglymosto

                                ALIGNMENTS


RESULT   1
P80913
ID   P80913      PRELIMINARY;      PRT;   130 AA.
AC   P80913;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   IG KAPPA CHAIN V REGION PRECURSOR.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   JANON K.;
RL   SUBMITTED (SEP-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; X79906; E269393; ALT_TERM.
KW   IMMUNOGLOBULIN V REGION; SIGNAL.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    130       IG KAPPA CHAIN V REGION.
SQ   SEQUENCE   130 AA;  14124 MW;  1DC6F7D9 CRC32;


  Query Match             34.4%;  Score 436;  DB 10;  Length 130;
  Best Local Similarity   80.0%;  Pred. No. 7.4e-25;
  Matches   84;  Conservative    9;  Mismatches   10;  Indels    2;  Gaps    1;

Qy        5 LTQSPAILSASPGEKVTMTCRATPSV--SYMHWYQQKPGSSPKPWIYTTSNLASGVPARF 62
            |||||||:||| ||:||||| |: ||  ||:|||||||||||| |||:||||||||||||
Db       26 LTQSPAIMSASLGERVTMTCTASSSVSSSYLHWYQQKPGSSPKLWIYSTSNLASGVPARF 85

Qy       63 SGGGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIK 107
            || |||||||||:| :|||||||||| |: | | |||||:|||||
Db       86 SGSGSGTSYSLTISSMEAEDAATYYCHQYHRFPHTFGGGTKLEIK 130


RESULT   2
Q15535
ID   Q15535      PRELIMINARY;      PRT;   100 AA.
AC   Q15535;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   V KAPPA (FRAGMENT).
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   KATO S., TACHIBANA K., TAKAYAMA N., KATAOKA H., YOSHIDA M.C.,
RA   TAKANO T.;
RL   SUBMITTED (SEP-1990) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; D90161; G1262585; -.
DR   PFAM; PF00047; ig; 1.
FT   NON_TER       1      1
FT   NON_TER     100    100
SQ   SEQUENCE   100 AA;  10871 MW;  06A1440D CRC32;

  Query Match             18.9%;  Score 239;  DB 2;  Length 100;
  Best Local Similarity   46.0%;  Pred. No. 8.1e-11;
  Matches   46;  Conservative   19;  Mismatches   29;  Indels    6;  Gaps    1;

Qy       2 DLQLTQSPAILSASPGEKVTMTCRATPSV------SYMHWYQQKPGSSPKPWIYTTSNLA 55
           |: :||:|  || :||:  :::|::  |:      :|:: | |||| ||:  ||  |:
Db       1 DVVMTQTPLSLSVTPGQPASISCKSIQSLLHSDGKTYLYXYLQKPGQSPQLLIYEVSSRF 60

Qy      56 SGVPARFSGGGSGTSYSLTVSRVEAEDAATYYCQQWSRSP 95
           |||| |||| |||| ::| :|||||||   ||| |    |.
Db      61 SGVPDRFSGSGSGTDFTLKISRVEAEDVGVYYCMQGIHLP 100


RESULT   3
O75737
ID   O75737      PRELIMINARY;      PRT;    88 AA.
AC   O75737;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN VARIABLE REGION (FRAGMENT).
GN   VH.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=INTESTINE;
RA   FISCHER M., KUEPPERS R.;
RT   "Human IgA and IgM secreting intestinal plasma cells carry heavily
RT   mutated VH region genes.";
RL   SUBMITTED (JUL-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AJ009535; E1311468; -.
FT   NON_TER       1      1
FT   NON_TER      88     88
SQ   SEQUENCE    88 AA;  10088 MW;  C76DC923 CRC32;

  Query Match             15.3%;  Score 193.5;  DB 2;  Length 88;
  Best Local Similarity   49.4%;  Pred. No. 1.3e-07;
  Matches   42;  Conservative   10;  Mismatches   20;  Indels   13;  Gaps    3;

Qy     156 WVKQSHGKNLEWIGLINPYNGDTNYNQKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYY 215
           ||:|: || ||||| |  | ||||||   ||: | |:||| :   : | |:|: ||||||
Db       5 WVRQAPGKGLEWIGEIY-YGGDTNYNPSLKGRLTLTLDKSMNQFSLILRSVTAADSAVYY 63

Qy     216 CAR---------RVTDWYF---DVW 228
           |||         |   :|:   |||
Db      64 CARGGGYHCIGGRCYRYYYNGVDVW 88


RESULT   4
O75731
ID   O75731      PRELIMINARY;      PRT;    79 AA.
AC   O75731;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN VARIABLE REGION (FRAGMENT).
GN   VH.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=INTESTINE;
RA   FISCHER M., KUEPPERS R.;
```

```
RT   "Human IgA and IgM secreting intestinal plasma cells carry heavily
RT   mutated VH region genes.";
RL   SUBMITTED (JUL-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AJ009528; E1311456; -.
FT   NON_TER       1      1
FT   NON_TER      79     79
SQ   SEQUENCE    79 AA;  8842 MW;  7C40774F CRC32;

  Query Match             14.4%;  Score 182.5;  DB 2;  Length 79;
  Best Local Similarity   41.0%;  Pred. No. 7.2e-07;
  Matches   32;  Conservative   18;  Mismatches   25;  Indels    3;  Gaps    1;

Qy     154 MNWVKQSHGKNLEWIGLINPYNGDTNYNQKFKGKATFTVDKSSSTAYMELLSLTSEDSAV 213
           | ||:|: || |||: :::   |  :| :   |: | : | | :| |::: ||  :|:||
Db       2 MTWVRQAPGKGLEWVSIVSAAAGSIHYAKSVMGRFTISRDNSKNTMYLQMNSLRVDDTAV 61

Qy     214 YYCARRVTDW---YFDVW 228
           ||||: | :|   ||| |
Db      62 YYCAKVVGNWGLGYFDNW 79


RESULT   5
O75739
ID   O75739      PRELIMINARY;      PRT;    78 AA.
AC   O75739;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN VARIABLE REGION (FRAGMENT).
GN   VH.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=INTESTINE;
RA   FISCHER M., KUEPPERS R.;
RT   "Human IgA and IgM secreting intestinal plasma cells carry heavily
RT   mutated VH region genes.";
RL   SUBMITTED (JUL-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AJ009538; E1311472; -.
FT   NON_TER       1      1
FT   NON_TER      78     78
SQ   SEQUENCE    78 AA;  9031 MW;  05D273B9 CRC32;

  Query Match             14.3%;  Score 181;  DB 2;  Length 78;
  Best Local Similarity   44.2%;  Pred. No. 9.1e-07;
  Matches   34;  Conservative   13;  Mismatches   26;  Indels    4;  Gaps    2;

Qy     155 NWVKQSHGKNLEWIGLINPYNGDTNYNQKFKGKATFTVDKSSSTAYMELLSLTSEDSAVY 214
           :|::|| || ||||| |  | |  |||   | : | ::  | :   : | |:|: |:|||
Db       3 SWIRQSPGKGLEWIGYIY-YTGSINYNPSLKSRVTISLGTSENQFSLNLTSVTAADTAVY 61

Qy     215 YCARRVTDW---YFDVW 228
           :|||  : |   ||| |
Db      62 FCARHSSSWPIYYFDNW 78


RESULT   6
O75723
ID   O75723      PRELIMINARY;      PRT;    78 AA.
AC   O75723;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN VARIABLE REGION (FRAGMENT).
GN   VH.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
```

```
RP    SEQUENCE FROM N.A.
RC    TISSUE=INTESTINE;
RA    FISCHER M., KUEPPERS R.;
RT    "Human IgA and IgM secreting intestinal plasma cells carry heavily
RT    mutated VH region genes.";
RL    SUBMITTED (JUL-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR    EMBL; AJ009520; E1311440; -.
FT    NON_TER       1      1
FT    NON_TER      78     78
SQ    SEQUENCE   78 AA;  9075 MW;  DFEDD569 CRC32;

  Query Match             14.1%;  Score 178.5;  DB 2;  Length 78;
  Best Local Similarity   43.6%;  Pred. No. 1.4e-06;
  Matches   34;  Conservative   14;  Mismatches   25;  Indels    5;  Gaps    1;

Qy    156 WVKQSHGKNLEWIGLINPYNGDTNYNQKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYY 215
          ||:|: || | |:  |:      ||    ||: | : | : :| |::: || :||:||||
Db      1 WVRQAPGKGLVWVXRISSDESRINYADSVKGRFTISRDNAKNTLYLQMNSLRAEDTAVYY 60

Qy    216 CARR-----VTDWYFDVW 228
          ||||      || :||||
Db     61 CARRXETTPATDRHFDVW 78


RESULT   7
O75729
ID    O75729       PRELIMINARY;      PRT;    82 AA.
AC    O75729;
DT    01-NOV-1998 (TREMBLREL. 08, CREATED)
DT    01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT    01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE    IG HEAVY CHAIN VARIABLE REGION (FRAGMENT).
GN    VH.
OS    HOMO SAPIENS (HUMAN).
OC    EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC    PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN    [1]
RP    SEQUENCE FROM N.A.
RC    TISSUE=INTESTINE;
RA    FISCHER M., KUEPPERS R.;
RT    "Human IgA and IgM secreting intestinal plasma cells carry heavily
RT    mutated VH region genes.";
RL    SUBMITTED (JUL-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR    EMBL; AJ009526; E1311452; -.
FT    NON_TER       1      1
FT    NON_TER      82     82
SQ    SEQUENCE   82 AA;  9396 MW;  9063B32A CRC32;

  Query Match             13.9%;  Score 176.5;  DB 2;  Length 82;
  Best Local Similarity   42.2%;  Pred. No. 2e-06;
  Matches   35;  Conservative   16;  Mismatches   25;  Indels    7;  Gaps    2;

Qy    152 HTMNWVKQSHGKNLEWIGLINPYNGDTNYNQKFKGKATFTVDKSSSTAYMELLSLTSEDS 211
          : |:||:|: || |||:  |:    || | :  ||: | : | | :| |::: :| :||:
Db      1 YAMSWVRQAPGKGLEWVSSIDGAGIDTYYAESVKGRFTISRDNSKNTLYLQMNTLRAEDT 60

Qy    212 AVYYC------ARRVTDWYFDVW 228
          |||||      |  | | ||| |
Db     61 AVYYCVKDGVSANSVWD-YFDYW 82


RESULT   8
O43234
ID    O43234       PRELIMINARY;      PRT;    97 AA.
AC    O43234;
DT    01-JUN-1998 (TREMBLREL. 06, CREATED)
DT    01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT    01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE    RHEUMATOID FACTOR RF-ET13 (FRAGMENT).
OS    HOMO SAPIENS (HUMAN).
OC    EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
```

```
OC    CATARRHINI; HOMINIDAE; HOMO.
RN    [1]
RP    SEQUENCE FROM N.A.
RA    BORRETZEN M., NATVIG J.B., THOMPSON K.M.;
RL    MOL. IMMUNOL. 0:0-0(1997).
DR    EMBL; AF035802; G2665882; -.
FT    NON_TER       1      1
FT    NON_TER      97     97
SQ    SEQUENCE   97 AA;  10748 MW;  EC95D20C CRC32;

  Query Match             13.9%;  Score 175.5;  DB 2;  Length 97;
  Best Local Similarity   35.1%;  Pred. No. 2.9e-06;
  Matches   34;  Conservative   23;  Mismatches   37;  Indels    3;  Gaps    2;

Qy    124 LQESGPEVVKPGGSMKISCKTSGYSFTGHTM--NWVKQSHGKNLEWIGLINPYNGDTNYN 181
          |:|||| :|||  :: ::|  ||:| :   |  :|::|  || :||:  |   | : :|:
Db      2 LKESGPALVKPTETLTLTCTVSGFSLSNRRMGVSWIRQPPGKAVEWLAHIFA-NDEKSYS 60

Qy    182 QKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCAR 218
             | : | : | | |   : : ::   |:| |||||
Db     61 TSLKSRLTISKDTSKSQVVLTMTNMDPMDTATYYCAR 97


RESULT   9
O75726
ID    O75726       PRELIMINARY;      PRT;    77 AA.
AC    O75726;
DT    01-NOV-1998 (TREMBLREL. 08, CREATED)
DT    01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT    01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE    IG HEAVY CHAIN VARIABLE REGION (FRAGMENT).
GN    VH.
OS    HOMO SAPIENS (HUMAN).
OC    EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC    PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN    [1]
RP    SEQUENCE FROM N.A.
RC    TISSUE=INTESTINE;
RA    FISCHER M., KUEPPERS R.;
RT    "Human IgA and IgM secreting intestinal plasma cells carry heavily
RT    mutated VH region genes.";
RL    SUBMITTED (JUL-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR    EMBL; AJ009523; E1311446; -.
FT    NON_TER       1      1
FT    NON_TER      77     77
SQ    SEQUENCE   77 AA;  9022 MW;  EB7B458D CRC32;

  Query Match             13.7%;  Score 174;  DB 2;  Length 77;
  Best Local Similarity   41.3%;  Pred. No. 2.9e-06;
  Matches   31;  Conservative   16;  Mismatches   28;  Indels    0;  Gaps    0;

Qy    154 MNWVKQSHGKNLEWIGLINPYNGDTNYNQKFKGKATFTVDKSSSTAYMELLSLTSEDSAV 213
          : ||:|: || |||:  ||   |:  |    ||: | : | :::: |::: || :||:||
Db      3 LTWVRQAPGKGLEWVATINEDGGEKYYVDSVKGRFTISRDDATNSLYLQMNSLRAEDTAV 62

Qy    214 YYCARRVTDWYFDVW 228
          ||| |   |  || |
Db     63 YYCVRDSYDRGFDFW 77


RESULT  10
O75724
ID    O75724       PRELIMINARY;      PRT;    85 AA.
AC    O75724;
DT    01-NOV-1998 (TREMBLREL. 08, CREATED)
DT    01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT    01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE    IG HEAVY CHAIN VARIABLE REGION (FRAGMENT).
GN    VH.
OS    HOMO SAPIENS (HUMAN).
OC    EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=INTESTINE;
RA   FISCHER M., KUEPPERS R.;
RT   "Human IgA and IgM secreting intestinal plasma cells carry heavily
RT   mutated VH region genes.";
RL   SUBMITTED (JUL-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AJ009521; E1311442; -.
FT   NON_TER       1      1
FT   NON_TER      85     85
SQ   SEQUENCE   85 AA;  9624 MW;  F1E1DCCC CRC32;


  Query Match             13.7%;  Score 173.5;  DB 2;  Length 85;
  Best Local Similarity   50.8%;  Pred. No. 3.5e-06;
  Matches   32;  Conservative   12;  Mismatches   18;  Indels    1;  Gaps    1;

Qy    156 WVKQSHGKNLEWIGLINPYNGDTNYNQKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYY 215
          |::|  || ||||| |  |:|:||||   | : | :|| | :   ::| |:|: |:||||
Db      6 WIRQPPGKGLEWIGFIY-YSGNTNYNPSLKSRLTISVDTSKNQFSLKLTSMTAADTAVYY 64

Qy    216 CAR 218
          |||
Db     65 CAR 67


RESULT  11
O75728
ID   O75728     PRELIMINARY;     PRT;    77 AA.
AC   O75728;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN VARIABLE REGION (FRAGMENT).
GN   VH.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=INTESTINE;
RA   FISCHER M., KUEPPERS R.;
RT   "Human IgA and IgM secreting intestinal plasma cells carry heavily
RT   mutated VH region genes.";
RL   SUBMITTED (JUL-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AJ009525; E1311450; -.
FT   NON_TER       1      1
FT   NON_TER      77     77
SQ   SEQUENCE   77 AA;  8925 MW;  EA14CFDF CRC32;


  Query Match             13.4%;  Score 170;  DB 2;  Length 77;
  Best Local Similarity   37.7%;  Pred. No. 5.6e-06;
  Matches   29;  Conservative   18;  Mismatches   30;  Indels    0;  Gaps    0;

Qy    152 HTMNWVKQSHGKNLEWIGLINPYNGDTNYNQKFKGKATFTVDKSSSTAYMELLSLTSEDS 211
          : |||::|: |: |:|:  :    | | |    ||: : : | | :| |::: || :||:
Db      1 YVMNWIRQAPGEGLQWVSTVTENGGVTFYPDSVKGRFSISRDNSRNTLYLQMNSLRAEDT 60

Qy    212 AVYYCARRVTDWYFDVW 228
          |:|||||      || |
Db     61 ALYYCARGKRSEGFDYW 77


RESULT  12
O75721
ID   O75721     PRELIMINARY;     PRT;    81 AA.
AC   O75721;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN VARIABLE REGION (FRAGMENT).
```

```
GN   VH.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=INTESTINE;
RA   FISCHER M., KUEPPERS R.;
RT   "Human IgA and IgM secreting intestinal plasma cells carry heavily
RT   mutated VH region genes.";
RL   SUBMITTED (JUL-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AJ009518; E1311436; -.
FT   NON_TER       1      1
FT   NON_TER      81     81
SQ   SEQUENCE   81 AA;  9466 MW;  ED58A5B2 CRC32;


  Query Match             13.4%;  Score 169.5;  DB 2;  Length 81;
  Best Local Similarity   38.8%;  Pred. No. 6.4e-06;
  Matches   31;  Conservative   19;  Mismatches   27;  Indels    3;  Gaps    1;

Qy    152 HTMNWVKQSHGKNLEWIGLINPYNGDTNYNQKFKGKATFTVDKSSSTAYMELLSLTSEDS 211
          ::||||:|: || |||: ||   :  | |    :|: | : | : :: |::: ||  :|:
Db      2 YSMNWVRQAPGKGLEWVSLIGTTDRATFYASSVRGRFTISRDDAKNSLYLQMTSLRDDDT 61

Qy    212 AVYYCARRVTDW---YFDVW 228
          ||||||| :  :   ||  |
Db     62 AVYYCARDLMIYPTEYFQDW 81


RESULT  13
O75742
ID   O75742     PRELIMINARY;     PRT;    76 AA.
AC   O75742;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN VARIABLE REGION (FRAGMENT).
GN   VH.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=INTESTINE;
RA   FISCHER M., KUEPPERS R.;
RT   "Human IgA and IgM secreting intestinal plasma cells carry heavily
RT   mutated VH region genes.";
RL   SUBMITTED (JUL-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AJ009542; E1311480; -.
FT   NON_TER       1      1
FT   NON_TER      76     76
SQ   SEQUENCE   76 AA;  8887 MW;  24B28E30 CRC32;


  Query Match             13.3%;  Score 168.5;  DB 2;  Length 76;
  Best Local Similarity   42.7%;  Pred. No. 7e-06;
  Matches   32;  Conservative   16;  Mismatches   26;  Indels    1;  Gaps    1;

Qy    154 MNWVKQSHGKNLEWIGLINPYNGDTNYNQKFKGKATFTVDKSSSTAYMELLSLTSEDSAV 213
          | ||:|: || |||: |    :  |:|    ||: | : | | || |::: ||  :|:||
Db      3 MFWVRQAPGKGLEWVALSERDDESTSYADSVKGRFTVSRDISKSTLYLQMSSLRVDDTAV 62

Qy    214 YYCARRVTDWYFDVW 228
          |:||:|   :|:| |
Db     63 YFCAKR-DGFYYDYW 76


RESULT  14
O75727
ID   O75727     PRELIMINARY;     PRT;    80 AA.
AC   O75727;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN VARIABLE REGION (FRAGMENT).
GN   VH.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=INTESTINE;
RA   FISCHER M., KUEPPERS R.;
RT   "Human IgA and IgM secreting intestinal plasma cells carry heavily
RT   mutated VH region genes.";
RL   SUBMITTED (JUL-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AJ009524; E1311448; -.
FT   NON_TER       1      1
FT   NON_TER      80     80
SQ   SEQUENCE   80 AA;  9351 MW;  63DE158A CRC32;

  Query Match             13.3%;  Score 168;  DB 2;  Length 80;
  Best Local Similarity   43.2%;  Pred. No. 8.1e-06;
  Matches   35;  Conservative   16;  Mismatches   22;  Indels    8;  Gaps    3;

Qy    154 MNWVKQSHGKNLEWIGLI----NPYNGDTNYNQKFKGKATFTVDKSSSTAYMELLSLTSE 209
          |:||:|: || |||:| |    | |   | |    ||: | : | | :: |::: || ::
Db      2 MDWVRQAPGKGLEWVGRIRKKANSYT--TEYAASVKGRFTISRDDSKNSLYLQMNSLKTD 59

Qy    210 DSAVYYCAR--RVTDWYFDVW 228
          |:|||||||    :  |||:|
Db     60 DTAVYYCARDHLRSAAYFDLW 80


RESULT  15
O75741
ID   O75741      PRELIMINARY;      PRT;    77 AA.
AC   O75741;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN VARIABLE REGION (FRAGMENT).
GN   VH.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=INTESTINE;
RA   FISCHER M., KUEPPERS R.;
RT   "Human IgA and IgM secreting intestinal plasma cells carry heavily
RT   mutated VH region genes.";
RL   SUBMITTED (JUL-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AJ009540; E1311476; -.
FT   NON_TER       1      1
FT   NON_TER      77     77
SQ   SEQUENCE   77 AA;  8734 MW;  1F7F9E8E CRC32;

  Query Match             13.1%;  Score 165.5;  DB 2;  Length 77;
  Best Local Similarity   44.7%;  Pred. No. 1.2e-05;
  Matches   34;  Conservative   13;  Mismatches   28;  Indels    1;  Gaps    1;

Qy    154 MNWVKQSHGKNLEWIGLINPYNGDTNYNQKFKGKATFTVDKSSSTAYMELLSLTSEDSAV 213
          |:||:|: || |||:  |:     | |    ||: | : | | :|  ::| |:|:||:||
Db      2 MSWVRQAPGKGLEWVSGISGSGDYTYYADSVKGRFTISRDNSKNTLSLQLNSVTAEDTAV 61

Qy    214 YYCARRVTDWY-FDVW 228
          |||||   |   || |
Db     62 YYCARHYYDSSPFDYW 77


Search completed: May 13, 1999, 10:05:19
```

Job time: 657 sec

```
                    GenCore version 4.5
           Copyright (c) 1993 - 1998  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         May 13, 1999, 09:37:42 ; Search time 23.53 Seconds
                                          (without alignments)
                                          203.715 Million cell updates/sec

Title:          US-08-704-178-3
Perfect score:  1267
Sequence:       1 MDLQLTQSPAILSASPGEKV..........RRVTDWYFDVWGAGTTVTVS 237

Scoring table:  BLOSUM62

Searched:       162890 seqs, 20225328 residues

Database :      A_Geneseq_34:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                            SUMMARIES
                 %
Result          Query
  No.   Score   Match Length DB  ID                    Description
-------------------------------------------------------------------------
    1    1267   100.0    237  1  R94020                Anti-erbB2 scFv. I
    2    1267   100.0    237  1  W15185                Single-chain anti-
    3    1256    99.1    236  1  R45442                Sequence of the si
    4   883.5    69.7    240  1  W25784                Anti-gp130 antibod
    5     863    68.1    244  1  R60780                Fv(GP-2) immunosup
    6     842    66.5    302  1  R60206                Bispecific CD3-L6F
    7   795.5    62.8    239  1  W15186                Single-chain anti-
    8   793.5    62.6    238  1  R45443                Sequence of the si
    9   784.5    61.9    243  1  R60781                Fv(GP-4) immunosup
   10   782.5    61.8    301  1  W37085                Anti-human SC sing
   11     748    59.0    272  1  R34672                26-10 sFv. Express
   12     746    58.9    251  1  P80152                Multifunctional pr
   13     745    58.8    239  1  W02191                18-2-3/TRY202' sin
   14     742    58.6    239  1  R43679                Single chain polyp
   15     742    58.6    239  1  R99649                Single chain bindi
   16     738    58.2    241  1  R06482                18-2-3-/TRY202'. S
   17   736.5    58.1    486  1  R37649                Sequence of divale
   18   736.5    58.1    262  1  R97380                CC49/212 SCA PLAP
   19   736.5    58.1    264  1  R97381                PLAP CC49/212 SCA
   20     734    57.9    241  1  R37645                Sequence of a mult
   21   731.5    57.7    250  1  R43678                Single chain polyp
   22   731.5    57.7    250  1  R99648                Single chain bindi
   23   731.5    57.7    250  1  W02190                TRY104b single cha
   24     730    57.6    271  1  R20185                Sequence of the sh
   25   726.5    57.3    246  1  R43677                Single chain polyp
   26   722.5    57.0    246  1  R99647                Single chain bindi
   27     722    57.0    473  1  W26646                Chimeric receptor
   28     722    57.0    514  1  W26647                Chimeric receptor
   29     722    57.0    403  1  W26648                Chimeric receptor
   30     722    57.0    651  1  W26649                Chimeric receptor
   31     722    57.0    692  1  W26650                Chimeric receptor
   32     722    57.0    382  1  W26651                Chimeric receptor
   33     715    56.4    238  1  R55864                4-4-20 VL / 217 /
   34   713.5    56.3    250  1  R06481                TRY104b. Single po
   35     710    56.0    238  1  R37650                Sequence of single
   36   705.5    55.7    248  1  R77616                Humanised CDR-graf
   37   704.5    55.6    242  1  R43680                Single chain polyp
   38   704.5    55.6    242  1  R99650                Single chain bindi
   39   704.5    55.6    242  1  W02192                18-2-3/TRY59 singl
   40   699.5    55.2    248  1  R77606                Murine 5G1.1M1 scF
   41   698.5    55.1    246  1  W02189                TRY59 single chain
   42   694.5    54.8    242  1  R06483                18-2-3-/TRY59. Sin
   43     693    54.7    272  1  W00557                Nematode salivary
```

```
   44   692.5    54.7    248  1  R77607                Humanised CDR-graf
   45     689    54.4    272  1  W43913                Mus musculus antib

                            ALIGNMENTS

RESULT   1
R94020
ID   R94020 standard; Protein; 237 AA.
AC   R94020;
DT   21-MAY-1996  (first entry)
DE   Anti-erbB2 scFv.
KW   Oncoprotein; erbB2; cell proliferation; tumour; cancer;
KW   intracellular antibody homologue; single chain antibody; scFv;
KW   gene therapy.
OS   Synthetic.
PN   WO9607321-A1.
PD   14-MAR-1996.
PF   23-AUG-1995; U10740.
PR   06-SEP-1994; US-301339.
PR   06-JUN-1995; US-468252.
PA   (UABR-) UAB RES FOUND.
PI   Curiel DT,  Deshane J;
DR   WPI; 96-171307/17.
DR   N-PSDB; T17728.
PT   Inhibition of proliferation or survival of, esp. malignant erbB2,
PT   cells - by introducing nucleic acid mol. encoding antibody homologue
PT   which is expressed and binds, pref. erbB2, protein intracellularly
PS   Disclosure; Page 29-30; 48pp; English.
CC   A nucleic acid comprises a first sequence encoding a signal peptide
CC   (R94019) linked to a second sequence (T17728) encoding a single
CC   chain Fv fragment (R94020) that binds a human erbB2 oncoprotein.
CC   The anti-erbB2 sFv portion is obtained by PCR using e23scFv
CC   plasmid as template.  The signal peptide directs the scFv to the
CC   endoplasmic reticulum.  The nucleic acid is incorporated into a
CC   plasmid or viral vector to facilitate expression of the scFv antibody
CC   homologue within e.g. an epithelial carcinoma cell.  Intracellular
CC   expression of the homologue inhibits surface expression of erbB2 and
CC   thereby inhibits cell proliferation and cell survival and decreases
CC   tumorigenicity.
SQ   Sequence   237 AA;


  Query Match             100.0%;  Score 1267;  DB 1;  Length 237;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-80;
  Matches  237;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 MDLQLTQSPAILSASPGEKVTMTCRATPSVSYMHWYQQKPGSSPKPWIYTTSNLASGVPA 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 MDLQLTQSPAILSASPGEKVTMTCRATPSVSYMHWYQQKPGSSPKPWIYTTSNLASGVPA 60

Qy       61 RFSGGGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIKGSTSGSGKSSEGK 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 RFSGGGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIKGSTSGSGKSSEGK 120

Qy      121 GVQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGDTNY 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 GVQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGDTNY 180

Qy      181 NQKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARRVTDWYFDVWGAGTTVTVS 237
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 NQKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARRVTDWYFDVWGAGTTVTVS 237


RESULT   2
W15185
ID   W15185 standard; Protein; 237 AA.
AC   W15185;
DT   05-JUN-1997  (first entry)
DE   Single-chain anti-erbB2 antibody e23(Fv).
KW   Single chain antibody; variable region; light chain; heavy chain;
KW   breast cancer; ovarian cancer; non-small cell lung carcinoma;
KW   immunodiagnosis; treatment; cytotoxic agent; erbB-2.
OS   Mus musculus.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   region          108. .121
FT                   /label= linker
PN   US5587458-A.
PD   24-DEC-1996.
PF   07-OCT-1991; 772270.
PR   07-OCT-1991; US-772270.
PR   30-JUN-1992; US-906555.
PR   14-MAY-1993; US-061092.
PA   (ARON-) ARONEX PHARM INC.
PI   Bird RE,  Kasprzyk PG,  King CR;
DR   WPI; 97-064831/06.
DR   N-PSDB; T65006.
PT   Single chain antibodies specific for erbB-2 protein, gp185  - with
PT   labels or cytotoxin, useful for detection and treatment of tumour
PT   cells expressing this protein
PS   Claim 2; Columns 25-28; 28pp; English.
CC   The present sequence represents a claimed single-chain antibody,
CC   designated e23(Fv), which binds to erbB-2. Monoclonal antibody e23
CC   was generated by immunising mice with N/erbB-2 cells overexpressing
CC   the gp185 protein, removing spleen cells and producing hybridomas
CC   by standard techniques. Messenger RNA coding for the anti-erbB-2
CC   monoclonal antibody was isolated and converted to cDNA. Regions
CC   coding for the heavy- and light- chain variable regions were then
CC   amplified by PCR and joined via a sequence encoding a peptide
CC   linker. The resulting single-chain antibody is useful for in vitro
CC   diagnosis of tumour cells which overexpress the erbB-2 gp185
CC   marker, e.g. breast, ovarian and non-small cell lung carcinomas,
CC   and, when coupled to a cytotoxic agent, to treat such tumours.
SQ   Sequence   237 AA;


  Query Match             100.0%;  Score 1267;  DB 1;  Length 237;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-80;
  Matches  237;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 MDLQLTQSPAILSASPGEKVTMTCRATPSVSYMHWYQQKPGSSPKPWIYTTSNLASGVPA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 MDLQLTQSPAILSASPGEKVTMTCRATPSVSYMHWYQQKPGSSPKPWIYTTSNLASGVPA 60

Qy       61 RFSGGGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIKGSTSGSGKSSEGK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 RFSGGGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIKGSTSGSGKSSEGK 120

Qy      121 GVQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGDTNY 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 GVQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGDTNY 180

Qy      181 NQKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARRVTDWYFDVWGAGTTVTVS 237
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 NQKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARRVTDWYFDVWGAGTTVTVS 237


RESULT   3
R45442
ID   R45442 standard; Protein; 236 AA.
AC   R45442;
DT   21-JUL-1994  (first entry)
DE   Sequence of the single chain anti-erbB2 antibody, Ab no.23.
KW   Single chain anti-erbB1 antibody; cancer therapy; prevention;
KW   monoclonal antibody.
OS   Synthetic.
PN   WO9400136-A.
PD   06-JAN-1994.
PF   21-OCT-1992; U08545.
PR   30-JUN-1992; US-906555.
PA   (MOLE-) MOLECULAR ONCOLOGY INC.
PI   Kasprzyk PG,  King CR;
DR   WPI; 94-025878/03.
DR   N-PSDB; Q55180.
```

```
PT   Treatment of malignancies over-expressing ERB-[2 - using at least
PT   2 monoclonal antibodies which recognise different epitopes on
PT   gp185
PS   Example; Fig 7; 37pp; English.
CC   The source of human erbB-2 protein for the prodn. of antibodies no.
CC   23 and 21 (Q55180/R45442; Q55181/R45443) is a NIH/3T3 cell
CC   engineered to express the human erbB-2 protein on its surface (N/
CC   erbB-2). Abs no. 21 and 23 are directed against the extracellular
CC   domain of gp185 erbB-2. Nude mice manipulated to produce rapidly
CC   growing tumours were used in a trial of the efficacy of the Abs. In
CC   animals given a combination of the 2 Abs, tumours completely
CC   regressed after 11 days.
SQ   Sequence   236 AA;

  Query Match             99.1%;  Score 1256;  DB 1;  Length 236;
  Best Local Similarity   99.2%;  Pred. No. 1.9e-79;
  Matches  234;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 DLQLTQSPAILSASPGEKVTMTCRATPSVSYMHWYQQKPGSSPKPWIYTTSNLASGVPAR 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 DLQLTQSPAILSASPGEKVTMTCRATPSVSYMHWYQQKPGSSPKPWIYTTSNLASGVPAR 60

Qy       62 FSGGGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIKGSTSGSGKSSEGKG 121
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 FSGGGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIKGSTSGSGKSSEGKG 120

Qy      122 VQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGDTNYN 181
              ||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 VQLQDSGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGDTNYN 180

Qy      182 QKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARRVTDWYFDVWGAGTTVTVS 237
              |||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db      181 QKFKGKATFTVDKSSSTAYMELLSLTSEDSALYYCARRVTDWYFDVWGAGTTVTVS 236


RESULT   4
W25784
ID   W25784 standard; Protein; 240 AA.
AC   W25784;
DT   22-DEC-1997  (first entry)
DE   Anti-gp130 antibody derived scFv.
KW   Single chain Fv; scFv; antibody; E. coli; anti-T3 antibody;
KW   anti-gp130 antibody; inclusion body; chaperonin.
OS   Synthetic.
PN   J09220092-A.
PD   26-AUG-1997.
PF   15-FEB-1996; 027622.
PR   15-FEB-1996; JP-027622.
PA   (TOYJ ) TOSOH CORP.
DR   WPI; 97-474306/44.
DR   N-PSDB; T91615.
PT   Producing single chain Fv antibody in Escherichia coli - by
PT   expression in an inclusion body, followed by protein folding or by
PT   co-expression with a chaperonin as a soluble fraction
PS   Claim 4; Page 6-7; 9pp; Japanese.
CC   The sequences given in W25783-84 represent single chain Fv (scFv)
CC   antibodies which are produced in E.coli.  The scFV's are derived
CC   from an anti-T3 antibody or an anti-gp130 antibody and are produced
CC   either by: (1) expression as an inclusion body, followed by folding
CC   (i.e. by denaturation and solubilisation) or (2) expression as a
CC   soluble fraction by co-expression with chaperonin.  Using the methods,
CC   scFv can be produced in E. coli, either in a soluble fraction or in
CC   inclusion bodies.  Like the parent double chain antibodies, the scFv
CC   polypeptides are useful in immunological diagnosis or for separation
CC   and purification, but they lessen the side effects caused by antibody
CC   constant regions.  The availability of scFv polypeptides contributes
CC   to fundamental research and development of diagnostic and therapeutic
CC   drugs.
SQ   Sequence   240 AA;

  Query Match             69.7%;  Score 883.5;  DB 1;  Length 240;
  Best Local Similarity   71.1%;  Pred. No. 6.5e-54;
  Matches  170;  Conservative   24;  Mismatches   42;  Indels    3;  Gaps    2;

Qy         2 DLQLTQSPAILSASPGEKVTMTCRATPSVSYMHWYQQKPGSSPKPWIYTTSNLASGVPAR 61
             :| ||||||::||||||||||||  : ||: |:||||||||||||||| ||||||||| |
Db         1 ELVLTQSPALISASPGEKVTMTCNVSSSVTSMYWYQQKPGSSPKPWIYLTSNLASGVPPR 60

Qy        62 FSGGGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIKGSTSGSGKS-SEGK 120
              || ||||:||||:| :|||||||||||||| :| ||| |:|||: |  || | |   | 
Db        61 SSGSGSGTTYSLTISSLEAEDAATYYCQQWSTNPLTFGAGTKLELGGGGSGGGGSGGGGS 120

Qy       121 GVQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGDTNY 180
              ||||:||||:|||| |:|| || |||:|| : |:||||||||:||||| || ::|   |
Db       121 DVQLQQSGPELVKPGASVKIPCKASGYTFTDYNMDWVKQSHGKSLEWIGDINSHSGGIIY 180

Qy       181 NQKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARR--VTDWYFDVWGAGTTVTVS 237
             ||||| ||  ||||||||||||| ||||||:|||||||   |   :|| || |||:|||
Db       181 NQKFKDKAILTVDKSSSTAYMELRSLTSEDTAVYYCARTYYVYGHFFDYWGQGTTLTVS 239


RESULT    5
R60780
ID   R60780 standard; Protein; 244 AA.
AC   R60780;
DT   09-MAY-1995  (first entry)
DE   Fv(GP-2) immunosuppressive.
KW   MAb; monoclonal antibody; hybridoma; interleukin-2; IL-2;
KW   Fv; antibody variable region; GP-2; Fv(GP-2);
KW   immunosuppressive.
OS   Mus sp.
PN   EP-621338-A.
PD   26-OCT-1994.
PF   21-APR-1994; 106257.
PR   21-APR-1993; JP-094491.
PR   07-MAR-1994; JP-036065.
PA   (AJIN ) AJINOMOTO KK.
PI   Hamura J,  Kanayama Y,  Nakazawa H,  Shimamura T,  Sugamura K;
PI   Takeshita T;
DR   WPI; 94-325948/41.
DR   N-PSDB; Q73678.
PT   Immunosuppressant polypeptide - has ability to block
PT   interleukin-2 response
PS   Claim 8; Page 26-27; 37pp; English.
CC   MAb capable of binding to the gamma chain of the IL-2 receptor, and
CC   thus of blocking the IL-2 response, is produced by mouse hybridoma
CC   line GP-2 (FERM BP-4641).  DNA encoding the variable region of
CC   this MAb was expressed in E. coli, yielding Fv(GP-2) with
CC   immumosuppressive activity.
SQ   Sequence   244 AA;

  Query Match             68.1%;  Score 863;  DB 1;  Length 244;
  Best Local Similarity   68.3%;  Pred. No. 1.7e-52;
  Matches  166;  Conservative   30;  Mismatches   41;  Indels    6;  Gaps    3;

Qy         1 MDLQLTQSPAILSASPGEKVTMTCRATPSV--SYMHWYQQKPGSSPKPWIYTTSNLASGV 58
             ||: |||||:|:||| ||:||||| |: ||  ||:|||||||||||| |||:||||||||
Db         1 MDILLTQSPSIMSASLGERVTMTCTASSSVSSSYLHWYQQKPGSSPKLWIYSTSNLASGV 60

Qy        59 PARFSGGGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIK-GSTSGSGKSS 117
             |||||| |||||||||:| :|||||||||| |: ||| ||| |:|||:|   :||||  |
Db        61 PARFSGSGSGTSYSLTISSMEAEDAATYYCHQYHRSPLTFGAGTKLELKVDKSSGSGSES 120

Qy       118 EGKGVQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGD 177
             :   |:|:||| |:|:|| |:|:||| |||:|| : |:|||| ||: ||||| | | :| 
Db       121 KSTQVKLEESGSELVRPGASVKLSCKASGYTFTSYWMHWVKQRHGQGLEWIGNIYPGSGS 180

Qy       178 TNYNQKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARRVTDWYF---DVWGAGTTV 234
             |||::||| ||| ||| ||||||| | |||||||||||| |   :| :   | || ||:|
Db       181 TNYDEKFKSKATLTVDTSSSTAYMHLSSLTSEDSAVYYCTRSSRNWVYYAMDYWGQGTSV 240

Qy       235 TVS 237
             |||
Db       241 TVS 243


RESULT    6
R60206
ID   R60206 standard; Protein; 302 AA.
AC   R60206;
DT   14-MAR-1995  (first entry)
DE   Bispecific CD3-L6FvIg antibody derivative.
KW   fusion protein; recombinant bispecific single chain antibody;
KW   helical peptide linker; anti-L6 antibody; tumour cell antigen;
KW   anti-CD3 antibody; variable region.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   peptide         1. .23
FT                   /label= L6_VL_leader
FT   region          24. .271
FT                   /label= CD3_VL-VH_fusion
FT   region          134. .148
FT                   /label= (Gly4Ser)3_linker
FT   region          272. .274
FT                   /label= hinge
FT   region          275. .302
FT                   /label= Fv_helical_linker
PN   EP-610046-A.
PD   10-AUG-1994.
PF   31-JAN-1994; 300692.
PR   01-FEB-1993; US-013420.
PR   13-SEP-1993; US-121054.
PA   (BRIM ) BRISTOL-MYERS SQUIBB CO.
PI   Bajorath J,  Fell PH,  Gilliland LK,  Hayden MS,  Ledbetter JA;
PI   Linsley PS;
DR   WPI; 94-250885/31.
DR   N-PSDB; Q81076.
PT   Expression vector encoding bispecific fusion protein - having
PT   binding domains for separate targets joined by helical peptide,
PT   useful e.g. for diagnosis and treatment
PS   Example 1; Fig 11 and Page 29-31; 50pp; English.
CC   The VL and VH sequences of the anti-CD3 hybridoma G19-4 were
CC   amplified by PCR methods. A gene fusion was constructed from the
CC   two amplified domains and a (Gly4Ser)3 linker. The amino
CC   terminus of the VL-VH fusion cassette was fused at the SalI site to
CC   the L6 light chain variable region leader peptide and the
CC   carboxy-terminus was fused directly to the hinge region of the Fc
CC   domain at the BclI site and/or to a short "helical" peptide linker
CC   to construct the bispecific CD3-L6FvIg antibody derivative. The
CC   variable regions for L6 were fused in frame to the opposite end of
CC   the helical linker (not included in R60206).
SQ   Sequence   302 AA;

  Query Match             66.5%;  Score 842;  DB 1;  Length 302;
  Best Local Similarity   65.1%;  Pred. No. 5.8e-51;
  Matches  162;  Conservative   30;  Mismatches   45;  Indels   12;  Gaps    3;

Qy         1 MDLQLTQSPAILSASPGEKVTMTCRATPSV-SYMHWYQQKPGSSPKPWIYTTSNLASGVP 59
             :|:|:||: : |||| |::||::|||:  : :|::||||||  : |  || || | ||||
Db        23 VDIQMTQTTSSLSASLGDRVTISCRASQDIRNYLNWYQQKPDGTVKLLIYYTSRLHSGVP 82

Qy        60 ARFSGGGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIKGSTSGSGKSSEG 119
             :|||| |||| ||||:: :: || |||:||| :  | |||||:||  |    | |    |
Db        83 SRFSGSGSGTDYSLTIANLQPEDIATYFCQQGNTLPWTFGGGTKLVTKRELGGGGSGGGG 142

Qy       120 KG-------VQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLIN 172
              |       ||||:||||:|||| || :||| |||||||: :||:|||||||||||||||
Db       143 SGGGGSIDEVQLQQSGPELVKPGASMTMSCKASGYSFTGYIVNWLKQSHGKNLEWIGLIN 202

Qy       173 PYNGDTNYNQKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARR----VTDWYFDVW 228
             || | | |||||||||| ||||||||||||||||||||||||||||      :|||||||
Db       203 PYKGLTTYNQKFKGKATLTVDKSSSTAYMELLSLTSEDSAVYYCARSGYYGDSDWYFDVW 262

Qy       229 GAGTTVTVS 237
             ||||| |||
Db    263 GAGTTCTVS 271

RESULT   7
W15186
ID   W15186 standard; Protein; 239 AA.
AC   W15186;
DT   05-JUN-1997  (first entry)
DE   Single-chain anti-erbB2 antibody e21(Fv).
KW   Single chain antibody; variable region; light chain; heavy chain;
KW   breast cancer; ovarian cancer; non-small cell lung carcinoma;
KW   immunodiagnosis; treatment; cytotoxic agent; erbB-2.
OS   Mus musculus.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   region          106. .119
FT                   /label= linker
PN   US5587458-A.
PD   24-DEC-1996.
PF   07-OCT-1991; 772270.
PR   07-OCT-1991; US-772270.
PR   30-JUN-1992; US-906555.
PR   14-MAY-1993; US-061092.
PA   (ARON-) ARONEX PHARM INC.
PI   Bird RE,  Kasprzyk PG,  King CR;
DR   WPI; 97-064831/06.
DR   N-PSDB; T65007.
PT   Single chain antibodies specific for erbB-2 protein, gp185  - with
PT   labels or cytotoxin, useful for detection and treatment of tumour
PT   cells expressing this protein
PS   Claim 2; Columns 27-30; 28pp; English.
CC   The present sequence represents a claimed single-chain antibody,
CC   designated e21(Fv), which binds to erbB-2. Monoclonal antibody e21
CC   was generated by immunising mice with N/erbB-2 cells overexpressing
CC   the gp185 protein, removing spleen cells and producing hybridomas
CC   by standard techniques. Messenger RNA coding for the anti-erbB-2
CC   monoclonal antibody was isolated and converted to cDNA. Regions
CC   coding for the heavy- and light- chain variable regions were then
CC   amplified by PCR and joined via a sequence encoding a peptide
CC   linker. The resulting single-chain antibody is useful for in vitro
CC   diagnosis of tumour cells which overexpress the erbB-2 gp185
CC   marker, e.g. breast, ovarian and non-small cell lung carcinomas,
CC   and, when coupled to a cytotoxic agent, to treat such tumours.
SQ   Sequence   239 AA;

  Query Match             62.8%;  Score 795.5;  DB 1;  Length 239;
  Best Local Similarity   64.3%;  Pred. No. 6.9e-48;
  Matches  155;  Conservative   32;  Mismatches   45;  Indels    9;  Gaps    2;

Qy      3 LQLTQSPAILSASPGEKVTMTCRATPSVSYMHWYQQKPGSSPKPWIYTTSNLASGVPARF 62
          :||||||||:|||||||||||| |: ||| |||||||  :||| |:| || |||||| ||
Db      1 MQLTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRF 60

Qy     63 SGGGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIKGSTSGSGKSSEGKGV 122
          || ||| |||||:| :|||||||||| | |  | ||| |:||||||||||||||||||||
Db     61 SGSGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIKGSTSGSGKSSEGKGV 120

Qy    123 QLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGDTNYNQ 182
          |||:|| |:|: |  :|:||| | ::   : ::||||   : ||||| |:| ||:| |:
Db    121 QLQQSGVELVRGGALVKLSCKASDFNIKDYYIHWVKQRPEQGLEWIGWIHPENGNTVYDP 180

Qy    183 KFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARRVTDWYFDV------WGAGTTVTV 236
          ||:|||: | | ||: ||::| ||||||:||||||   : :|:        || ||:|||
Db    181 KFQGKASITADTSSNAAYLQLSSLTSEDTAVYYCA---SYYYSAYYAMYYWGQGTSVTV 237

Qy    237 S 237
          |
Db    238 S 238

RESULT   8
R45443.
```

```
ID   R45443 standard; Protein; 238 AA.
AC   R45443;
DT   21-JUL-1994  (first entry)
DE   Sequence of the single chain anti-erbB2 antibody, Ab no.21.
KW   Single chain anti-erbB1 antibody; cancer therapy; prevention;
KW   monoclonal antibody.
OS   Synthetic.
PN   WO9400136-A.
PD   06-JAN-1994.
PF   21-OCT-1992; U08545.
PR   30-JUN-1992; US-906555.
PA   (MOLE-) MOLECULAR ONCOLOGY INC.
PI   Kasprzyk PG,  King CR;
DR   WPI; 94-025878/03.
DR   N-PSDB; Q55181.
PT   Treatment of malignancies over-expressing ERB-[2 - using at least
PT   2 monoclonal antibodies which recognise different epitopes on
PT   gp185
PS   Example; Fig 8; 37pp; English.
CC   The source of human erbB-2 protein for the prodn. of antibodies no.
CC   23 and 21 (Q55180/R45442; Q55181/R45443) is a NIH/3T3 cell
CC   engineered to express the human erbB-2 protein on its surface (N/
CC   erbB-2). Abs no. 21 and 23 are directed against the extracellular
CC   domain of gp185 erbB-2. Nude mice manipulated to produce rapidly
CC   growing tumours were used in a trial of the efficacy of the Abs. In
CC   animals given a combination of the 2 Abs, tumours completely
CC   regressed after 11 days.
SQ   Sequence   238 AA;

  Query Match             62.6%;  Score 793.5;  DB 1;  Length 238;
  Best Local Similarity   64.6%;  Pred. No. 9.5e-48;
  Matches  155;  Conservative   31;  Mismatches   45;  Indels    9;  Gaps    2;

Qy      4 QLTQSPAILSASPGEKVTMTCRATPSVSYMHWYQQKPGSSPKPWIYTTSNLASGVPARFS 63
          ||||||||:|||||||||||| |: ||| |||||||  :||| |:| || |||||| |||
Db      1 QLTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRFS 60

Qy     64 GGGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIKGSTSGSGKSSEGKGVQ 123
          | ||| |||||:| :|||||||||| | |  | ||| |:|||||||||||||||||||||
Db     61 GSGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIKGSTSGSGKSSEGKGVQ 120

Qy    124 LQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGDTNYNQK 183
          ||:|| |:|: |  :|:||| | ::   : ::||||   : ||||| |:| ||:| |: |
Db    121 LQQSGVELVRGGALVKLSCKASDFNIKDYYIHWVKQRPEQGLEWIGWIHPENGNTVYDPK 180

Qy    184 FKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARRVTDWYFDV------WGAGTTVTVS 237
          |:|||: | | ||: ||::| ||||||:||||||   : :|:        || ||:||||
Db    181 FQGKASITADTSSNAAYLQLSSLTSEDTAVYYCA---SYYYSAYYAMYYWGQGTSVTVS 237

RESULT   9
R60781
ID   R60781 standard; Protein; 243 AA.
AC   R60781;
DT   09-MAY-1995  (first entry)
DE   Fv(GP-4) immunosuppressive.
KW   MAb; monoclonal antibody; hybridoma; interleukin-2; IL-2;
KW   Fv; antibody variable region; GP-4; Fv(GP-4);
KW   immunosuppressive.
OS   Mus sp.
PN   EP-621338-A.
PD   26-OCT-1994.
PF   21-APR-1994; 106257.
PR   21-APR-1993; JP-094491.
PR   07-MAR-1994; JP-036065.
PA   (AJIN ) AJINOMOTO KK.
PI   Hamura J,  Kanayama Y,  Nakazawa H,  Shimamura T,  Sugamura K;
PI   Takeshita T;
DR   WPI; 94-325948/41.
DR   N-PSDB; Q73679.
PT   Immunosuppressant polypeptide - has ability to block
PT   interleukin-2 response
```

```
PS   Claim 9; Page 27-28; 37pp; English.
CC   MAb capable of binding to the gamma chain of the IL-2 receptor, and
CC   thus of blocking the IL-2 response, is produced by mouse hybridoma
CC   line GP-4 (FERM BP-4640).  DNA encoding the variable region of
CC   this MAb was expressed in E. coli, yielding Fv(GP-4) with
CC   immumosuppressive activity.
SQ   Sequence   243 AA;

  Query Match             61.9%;  Score 784.5;  DB 1;  Length 243;
  Best Local Similarity   63.5%;  Pred. No. 4e-47;
  Matches  155;  Conservative   32;  Mismatches   48;  Indels    9;  Gaps    4;

Qy      1 MDLQLTQSPAILSASPGEKVTMTCRATPSV-SYMHWYQQKPGSSPKPWIYTTSNLASGVP 59
          ||: |||||| |||| || ||:||||: :: :|: ||||| | ||:  :|    || |||
Db      1 MDILLTQSPASLSASVGETVTITCRASGNIHNYLAWYQQKQGKSPQLLVYNAKTLADGVP 60

Qy     60 ARFSGGGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIK-GSTSGSGKSSE 118
          :|||| |||| ||| :: :: ||  :|||| :  :| |||||:|||:|   :||||  |:
Db     61 SRFSGSGSGTQYSLKINSLQPEDFGSYYCQHFWSTPWTFGGGTKLELKVEKSSGSGSESK 120

Qy    119 GKGVQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGDT 178
             |:|:|||||:|||| |:||||| |||||||: |:|||||| |:||||| |||||| |
Db    121 STQVKLEESGPELVKPGASVKISCKASGYSFTGYYMHWVKQSHVKSLEWIGRINPYNGAT 180

Qy    179 NYNQKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARRVTDWY-----FDVWGAGTT 233
          :||| || ||: ||||||||||||| ||||||||||||||   ::|     |  ||  ||
Db    181 SYNQNFKDKASLTVDKSSSTAYMELHSLTSEDSAVYYCARE--NYYGSSYGFAYWGQGTL 238

Qy    234 VTVS 237
          ||||
Db    239 VTVS 242


RESULT  10
W37085
ID   W37085 standard; Protein; 301 AA.
AC   W37085;
DT   14-JUL-1998  (first entry)
DE   Anti-human SC single chain Fv/protamine fusion protein.
KW   Fusion protein; SC single chain Fv/protamine fusion protein; SECR;
KW   exogenous gene; serpin enzyme complex receptor; gene therapy;
KW   target binding moiety.
OS   Homo sapiens.
OS   Mus sp.
PN   WO9746100-A1.
PD   11-DEC-1997.
PF   03-JUN-1997;  U09858.
PR   03-JUN-1996; US-656906.
PA   (UYCA-) UNIV CASE WESTERN RESERVE.
PI   Davis PB, Ferkol TW, Ziady A;
DR   WPI; 98-041783/04.
DR   N-PSDB; V00611.
PT   Delivering compacted exogenous nucleic acid to cells  - by targeting
PT   the serpin enzyme complex receptor, used in gene therapy
PS   Example 9; Pages 120-121; 158pp; English.
CC   This represents an anti-human SC single chain Fv/protamine fusion protein
CC   sequence containing a target binding moiety capable of binding to a
CC   serpin enzyme complex receptor (SECR), and a nucleic acid binding moiety.
CC   This can be used in a method for delivering an oligonucleotide to a
CC   mammalian cell. The method comprises conjugating the target binding
CC   moiety to a nucleic acid binding moiety to form a carrier and coupling
CC   the carrier to an expression vector encoding one or more gene products
CC   to form a pharmaceutical composition. A mammalian cell having on its
CC   surface SECR, is contacted with the pharmaceutical composition under
CC   conditions allowing binding to the receptor resulting in delivery of the
CC   pharmaceutical composition to the interior of the cell. The composition
CC   and method are used for the introduction of exogenous genetic material
CC   into target host cells expressing SECR on their surface. The nucleic acid
CC   may encode a functional wild-type or mutant gene or may be an antisense
CC   sequence or other nucleic acid having a therapeutic effect. The fusion
CC   protein may comprise a protein portion having therapeutic properties,
CC   e.g. enzymatic activity, cytokine activity and antibiotic activity which
```

```
CC   is delivered to a cell surface via the SECR binding moiety. The nucleic
CC   acid can be compacted at high concentrations with the carrier molecule at
CC   a critical salt concentration. The condensation of such complexes
CC   provides structural features to the DNA/cationic lipid complex that
CC   prolong in vivo expression.
SQ   Sequence   301 AA;

  Query Match             61.8%;  Score 782.5;  DB 1;  Length 301;
  Best Local Similarity   62.4%;  Pred. No. 6.9e-47;
  Matches  156;  Conservative   28;  Mismatches   49;  Indels   17;  Gaps    4;

Qy      1 MDLQLTQSPAILSASPGEKVTMTCRATPSV-----SYMHWYQQKPGSSPKPWIYTTSNLA 55
          ||: |||||| |: | |:: |::|||: ||     |:|:|:|||||  ||  ||  ||
Db      1 MDIVLTQSPASLAVSLGQRATISCRASESVDNYAISFMNWFQQKPGQPPKLLIYAASNQG 60

Qy     56 SGVPARFSGGGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIK-GSTSGSG 114
          |||| |||| |||| :|| :  :| :| | |:|||    | |||||:||||| |  || |
Db     61 SGVPGRFSGSGSGTDFSLNIHPMEEDDTAMYFCQQSKAVPYTFGGGTKLEIKGGGGSGGG 120

Qy    115 KSSEGKGVQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPY 174
           |  |  ||||:|||::|||| |:|||||||||:|  :||:||||||||:||||| |||
Db    121 GSGGGSEVQLQQSGPDLVKPGASVKISCKTSGYTFIEYTMHWVKQSHGKSLEWIGGINPN 180

Qy    175 NGDTNYNQKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARRVTDWY-------FDV 227
          || |:|||||||||| ||||||:||||||  ||||||||| |||    :|        |
Db    181 NGGTSYNQKFKGKATLTVDKSSTTAYMELRGLTSEDSAVYSCAR----YYRYDVLSAMDY 236

Qy    228 WGAGTTVTVS 237
          || |||:|||
Db    237 WGQGTTLTVS 246


RESULT  11
R34672
ID   R34672 standard; Protein; 272 AA.
AC   R34672;
DT   25-AUG-1993  (first entry)
DE   26-10 sFv.
KW   Heavy; light; variable; VH; VL; region; antidigitoxin; monoclonal;
KW   antibody; MAb; 26-10; Fv; fragment; antigen; binding site; linker;
KW   expression-secretion system; T7 promoter; signal peptide; molecule;
KW   polymerase chain reaction; PCR; single chain.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   peptide         1. .25
FT                   /note= "Signal peptide"
FT   region          26. .138
FT                   /note= "26-10 VL sequence"
FT   peptide         139. .153
FT                   /label= Linker peptide
FT   region          154. .272
FT                   /note= "26-10 VH sequence"
PN   WO9308300-A.
PD   29-APR-1993.
PF   16-OCT-1992;  U08881.
PR   18-OCT-1991; US-777709.
PA   (UYCA-) UNIV CALGARY.
PI   Anthony JG,  Ng SC,  Wong S;
DR   WPI; 93-152491/18.
DR   N-PSDB; Q41069.
PT   Expression-secretion vectors - for prodn. of biologically active
PT   antibody Fv fragments or single chain Fv molecules
PS   Disclosure; Fig 6; 51pp; English.
CC   This sequence represents a single chain Fv molecule which contains
CC   the heavy and light variable portions (VH and VL) of antidigitoxin
CC   monoclonal antibody (MAb) 26-10 (see also R34670-71), a signal
CC   sequence (see R38527-28) and a linker between the two variable
CC   regions.  The DNA sequence encoding this protein was used in the
CC   construction of an expression-secretion system for the production
CC   of biologically active Fv fragments.  The system also contains a
CC   DNA sequence encoding the T7 promoter.  The secretion-expression
CC   vector was produced by polymerase chain reaction Fv fragments
CC   comprising only the VH or VL regions, and single chain molecules may
CC   be used to specifically bind one or more of the same antigens as the
CC   full length antibody from which they are derived.
SQ   Sequence   272 AA;

  Query Match             59.0%;  Score 748;  DB 1;  Length 272;
  Best Local Similarity   60.2%;  Pred. No. 1.4e-44;
  Matches  148;  Conservative   33;  Mismatches   55;  Indels   10;  Gaps    4;

Qy          2 DLQLTQSPAILSASPGEKVTMTCRATPSV------SYMHWYQQKPGSSPKPWIYTTSNLA 55
              |: :||:|  |  | |:: :::||:: |:      :|::|| || | |||  ||  ||
Db         26 DVVMTQTPLSLPVSLGDQASISCRSSQSLVHSNGNTYLNWYLQKAGQSPKLLIYKVSNRF 85

Qy         56 SGVPARFSGGGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIK--GSTS-G 112
              |||| |||| |||| ::| :|||||||   |:| | :   ||||||:|||||  || | |
Db         86 SGVPDRFSGSGSGTDFTLKISRVEAEDLGIYFCSQTTHVPPTFGGGTKLEIKRGGSGSGG 145

Qy        113 SGKSSEGKGVQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLIN 172
              ||    |  ||||:||||:|||| |:::|||:||| ||   ||||:|||||:|::|| |:
Db        146 SGSGGSGSEVQLQQSGPELVKPGASVRMSCKSSGYIFTDFYMNWVRQSHGKSLDYIGYIS 205

Qy        173 PYNGDTNYNQKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCA-RRVTDWYFDVWGAG 231
              ||:| | |||||||||| ||||||||||||| |||||||||||||      |  | || |
Db        206 PYSGVTGYNQKFKGKATLTVDKSSSTAYMELRSLTSEDSAVYYCAGSSGNKWAMDYWGHG 265

Qy        232 TTVTVS 237
               :||||
Db        266 ASVTVS 271


RESULT  12
P80152
ID   P80152 standard; protein; 251 AA.
AC   P80152;
DT   20-OCT-1990  (first entry)
DE   Multifunctional protein.
KW   Multifunctional protein; biosynthetic antibody binding site;
KW   murine monoclonal 26-10.
PN   WO8809344-A.
PD   01-DEC-1988.
PF   19-MAY-1988; U01737.
PR   21-MAY-1987; US-052800.
PA   (CREA-) Creative Biomolecules Inc.
PI   Huston JS, Oppermann H;
DR   WPI; 88-353928/49.
DR   N-PSDB; N80178.
PT   Recombinant multifunctional protein - having antibody binding site and a
PT   sequence for biological activity, ion sequestering or binding to a solid
PT   support.
PS   Disclosure; 15pp; English.
CC   The multifunctional protein comprises a single chain biosynthetic Ab
CC   binding site (comprising sequences mimicking the light chain and heavy
CC   chain variable regions of murine monoclonal 26-10) with the specificity
CC   of murine monoclonal 26-10, linked through a spacer to the FB fragment of
CC   protein A, here fused as a leader, and constituting a binding site for Fc
CC   The spacer contains the 11 C-terminal amino acids of the FB followed by
CC   Asp-Pro ( a dilute acid cleavage site).  The construct contains
CC   binding sites for both Fc and digoxin, and its structure can be
CC   summarised as: FB-Asp-Pro-VL-(Gly4-Ser)3-VH.  The protein can be used for
CC   specific binding assays, affinity purificn., biocatalysts, targetting,
CC   imaging an immunological treatment of oncogenic and infectious diseases.
CC   It offers fewer cleavage sites to circulating proteolytic enzymes, and
CC   improved stability.  It reaches target tissues rapidly and is quickly
CC   cleared from the body.  It also has reduced immunogenicity and its design
CC   facilitates binding to other moieties in drug targetting and imaging
CC   applications.
SQ   Sequence   251 AA;

  Query Match             58.9%;  Score 746;  DB 1;  Length 251;
  Best Local Similarity   58.5%;  Pred. No. 1.8e-44;
  Matches  145;  Conservative   34;  Mismatches   57;  Indels   12;  Gaps    3;
```

```
Qy          2 DLQLTQSPAILSASPGEKVTMTCRATPSV------SYMHWYQQKPGSSPKPWIYTTSNLA 55
              |: :||:|  |  | |:: :::||:: |:      :|::|| || | |||  ||  ||
Db          3 DVVMTQTPLSLPVSLGDQASISCRSSQSLVHSNGNTYLNWYLQKAGQSPKLLIYKVSNRF 62

Qy         56 SGVPARFSGGGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIK-----GST 110
              |||| |||| |||| ::| :|||||||   |:| | :   ||||||:|||||     | :
Db         63 SGVPDRFSGSGSGTDFTLKISRVEAEDLGIYFCSQTTHVPPTFGGGTKLEIKRGSGGGGS 122

Qy        111 SGSGKSSEGKGVQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGL 170
               | |    |  ||||:||||:|||| |:::|||:||| ||   ||||:|||||:|::||
Db        123 GGGGSGGGGSEVQLQQSGPELVKPGASVRMSCKSSGYIFTDFYMNWVRQSHGKSLDYIGY 182

Qy        171 INPYNGDTNYNQKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCA-RRVTDWYFDVWG 229
              |:||:| | |||||||||| ||||||||||||| |||||||||||||      |  | ||
Db        183 ISPYSGVTGYNQKFKGKATLTVDKSSSTAYMELRSLTSEDSAVYYCAGSSGNKWAMDYWG 242

Qy        230 AGTTVTVS 237
               | :||||
Db        243 HGASVTVS 250


RESULT  13
W02191
ID   W02191 standard; Protein; 239 AA.
AC   W02191;
DT   13-NOV-1996  (first entry)
DE   18-2-3/TRY202' single chain binding protein.
KW   Antibody engineering; monoclonal antibody; MAb;
KW   single chain antibody; immunoassay; fluorescein; 18-2-3/TRY202';
KW   single chain binding protein.
OS   Chimeric Mus musculus;
OS   Chimeric synthetic.
PN   US5518889-A.
PD   21-MAY-1996.
PF   02-SEP-1986; 902971.
PR   02-SEP-1986; US-902971.
PR   02-SEP-1987; US-092110.
PR   19-JAN-1989; US-299617.
PR   25-APR-1990; US-512910.
PR   01-APR-1993; US-040440.
PR   06-JUN-1995; US-468988.
PA   (ENZO-) ENZON LABS INC.
PI   Bird RE,  Hardman K,  Ladner RC;
DR   WPI; 96-259060/26.
DR   N-PSDB; T36463.
PT   Immunoassay using single chain antigen binding mol. - as replacement
PT   for labelled or immobilised antibody, are less immunogenic, easier
PT   to engineer, more stable and less expensive
PS   Example 9; Fig 39A-B; 78pp; English.
CC   Computer-designed construct 18-2-3/TRY202' (W02191), encoded by
CC   the DNA sequence given in T36463, comprises the variable regions
CC   of the light and heavy chains of anti-fluorescein monoclonal
CC   antibody (Mab) 18-2-3 linked by a peptide designed to fit into a
CC   groove on the backside of the variable domain structure.  The DNA
CC   construct was inserted into vector pGX3703 and introduced into E.
CC   coli.  18-2-3/TRY202' was produced as a single chain molecule.  It
CC   exhibited biological binding activity equivalent in specifically
CC   and affinity to that of the original MAb.
SQ   Sequence   239 AA;

  Query Match             58.8%;  Score 745;  DB 1;  Length 239;
  Best Local Similarity   59.5%;  Pred. No. 2e-44;
  Matches  144;  Conservative   40;  Mismatches   50;  Indels    8;  Gaps    4;

Qy          1 MDLQLTQSPAILSASPGEKVTMTCRATPSV--SYMHWYQQKPGSSPKPWIYTTSNLASGV 58
              |:  |||||||:||||||||||||||: ||  ||:|||||| |:||| |:| ||||||||
Db          1 MENVLTQSPAIMSASPGEKVTMTCRASSSVSSSYLHWYQQKSGASPKLWVYGTSNLASGV 60

Qy         59 PARFSGGGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIKGSTSGSGKSSE 118
              ||||||||||||||||:| |||||||||||||:|  | ||| |:|||::| :||||  ||
Db         61 PARFSGGGSGTSYSLTISSVEAEDAATYYCQQYSGYPLTFGAGTKLELEGKSSGSG--SE 118
```

```
Qy    119 GKGVQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGDT 178
           |  ||:|||| :| |  |: |:|  ||:| | : ::||:|  || |||:|:|    |:|
Db    119 SKSTQLKESGPVLVAPSQSLSITCTVSGFSLTNYGVHWVRQPPGKGLEWLGVIWA-GGNT 177

Qy    179 NYNQKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARRVTDWYF---DVWGAGTTVT 235
          |||     : : : | | |  :::: || ::|:|:||||:|:   ::   | || ||:||
Db    178 NYNSALMSRLSISKDNSKSQVFLKMNSLQNDDTAIYYCAKRLGRIFYYAMDYWGQGTSVT 237

Qy    236 VS 237
          ||
Db    238 VS 239


RESULT  14
R43679
ID   R43679 standard; Protein; 239 AA.
AC   R43679;
DT   23-MAY-1994  (first entry)
DE   Single chain polypeptide with affinity for fluorescein.
KW   Monoclonal antibody; MAb; affinity; binding; antigen; diagnostics;
KW   therapy; imaging; purification; biosensors.
OS   Synthetic.
PN   US5260203-A.
PD   09-NOV-1993.
PF   02-SEP-1986; 902971.
PR   02-SEP-1986; US-902971.
PR   02-SEP-1987; US-092110.
PR   19-JAN-1989; US-299617.
PR   25-APR-1990; US-512910.
PA   (ENZO-) ENZON INC.
PI   Bird RE,  Hardman K,  Ladner RC;
DR   WPI; 93-367875/46.
DR   N-PSDB; Q51540.
PT   Single chain poly:peptide for binding antigen - comprising light
PT   and heavy chain antigen binding portions linked by peptide linker
PS   Example 9; Figure 39; 78pp; English.
CC   The single chain polypeptide is derived from the mature light and
CC   heavy chains of a monoclonal antibody (MAb) and has affinity
CC   for a given antigen (Fluorescein).  It comprises a first
CC   polypeptide comprising the antigen binding portion of of the light
CC   chain variable region of an antibody and a second polypeptide
CC   comprising the antigen binding portion of the heavy chain variable
CC   region of an antibody and at least one peptide linker linking the
CC   first and second polypeptide chains.  The resulting single chain
CC   polypeptide can be used in diagnostics, therapy
CC   (in vivo and in vitro), imaging, purifications and biosensors.
CC   This particular single chain binding molecule was designated
CC   18-2-3/TRY202' and contains one linker peptide.
SQ   Sequence   239 AA;

  Query Match             58.6%;  Score 742;  DB 1;  Length 239;
  Best Local Similarity   59.5%;  Pred. No. 3.2e-44;
  Matches  144;  Conservative   39;  Mismatches   51;  Indels    8;  Gaps    4;

Qy      1 MDLQLTQSPAILSASPGEKVTMTCRATPSV--SYMHWYQQKPGSSPKPWIYTTSNLASGV 58
          |:  |||||||:||||||||||||||: ||  ||:|||||| |:||| |:| ||||||||
Db      1 MENVLTQSPAIMSASPGEKVTMTCRASSSVSSSYLHWYQQKSGASPKLWVYGTSNLASGV 60

Qy     59 PARFSGGGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIKGSTSGSGKSSE 118
          ||||||||||||||||:| |||||||||||||:|  | ||| |:|||::| :||||  ||
Db     61 PARFSGGGSGTSYSLTISSVEAEDAATYYCQQYSGYPLTFGAGTKLELEGKSSGSG--SE 118

Qy    119 GKGVQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGDT 178
           |  ||:|||| :| |  |: |:|  ||:| | : ::||:|  || |||:|:|    |:|
Db    119 SKSTQLKESGPVLVAPSQSLSITCTVSGFSLTNYGVHWVRQPPGKGLEWLGVIWA-GGNT 177

Qy    179 NYNQKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARRVTDWYF---DVWGAGTTVT 235
          |||     : : : | | |  :::: ||  :|:|:||||:|:   ::   | || ||:||
Db    178 NYNSALMSRLSISKDNSKSQVFLKMNSLQIDDTAIYYCAKRLGRIFYYAMDYWGQGTSVT 237

Qy    236 VS 237
          ||
Db    238 VS 239


RESULT  15
R99649
ID   R99649 standard; Protein; 239 AA.
AC   R99649;
DT   11-OCT-1996  (first entry)
DE   Single chain binding molecule 18-2-3/TRY202'.
KW   Antibody engineering; single polypeptide chain binding molecule;
KW   single chain antibody; SCA; heavy chain; light chain;
KW   monoclonal antibody; MAb; immunoaffinity purification;
KW   18-2-3/TRY202'.
OS   Chimeric Mus sp;
OS   Chimeric synthetic.
PN   US5534621-A.
PD   09-JUL-1996.
PF   02-SEP-1986; 902971.
PR   02-SEP-1986; US-902971.
PR   02-SEP-1987; US-092110.
PR   19-JAN-1989; US-299617.
PR   25-APR-1990; US-512910.
PR   01-APR-1993; US-040440.
PR   06-JUN-1995; US-468992.
PA   (ENZO-) ENZON LABS INC.
PI   Bird RE,  Hardman K,  Ladner RC;
DR   WPI; 96-333309/33.
DR   N-PSDB; T13739.
PT   Immuno:purificn. using single binding chain molecule including
PT   antigen-binding parts of antibody light and heavy chain variable
PT   regions connected by a linker - is smaller, stabler and less
PT   expensive than complete antibodies
PS   Example 9; Fig 39; 78pp; English.
CC   A DNA construct (T13739) codes for single chain binding molecule
CC   18-2-3/TRY202' (R99649), in which VL and VH regions of anti-
CC   fluorescein monoclonal antibody 18-2-3 are joined by a peptide
CC   linker composed primarily of alternating Gly and Ser residues, with
CC   Glu and Lys residues to enhance solubility.  The construct was
CC   expressed in E.coli.  Soluble, folded chain antibody (SCA) was
CC   obtd. which was capable of exhibiting a biological binding activity
CC   equivalent in specificity and affinity to that of a monoclonal
CC   antibody.  SCAs (see also R99645-48 and T99650-51) have the
CC   advantages of smaller size, greater stability and reduced cost.
SQ   Sequence   239 AA;

  Query Match             58.6%;  Score 742;  DB 1;  Length 239;
  Best Local Similarity   59.5%;  Pred. No. 3.2e-44;
  Matches  144;  Conservative   39;  Mismatches   51;  Indels    8;  Gaps    4;

Qy      1 MDLQLTQSPAILSASPGEKVTMTCRATPSV--SYMHWYQQKPGSSPKPWIYTTSNLASGV 58
          |:  |||||||:||||||||||||||: ||  ||:|||||| |:||| |:| ||||||||
Db      1 MENVLTQSPAIMSASPGEKVTMTCRASSSVSSSYLHWYQQKSGASPKLWVYGTSNLASGV 60

Qy     59 PARFSGGGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIKGSTSGSGKSSE 118
          ||||||||||||||||:| |||||||||||||:|  | ||| |:|||::| :||||  ||
Db     61 PARFSGGGSGTSYSLTISSVEAEDAATYYCQQYSGYPLTFGAGTKLELEGKSSGSG--SE 118

Qy    119 GKGVQLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGDT 178
           |  ||:|||| :| |  |: |:|  ||:| | : ::||:|  || |||:|:|    |:|
Db    119 SKSTQLKESGPVLVAPSQSLSITCTVSGFSLTNYGVHWVRQPPGKGLEWLGVIWA-GGNT 177

Qy    179 NYNQKFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARRVTDWYF---DVWGAGTTVT 235
          |||     : : : | | |  :::: ||  :|:|:||||:|:   ::   | || ||:||
Db    178 NYNSALMSRLSISKDNSKSQVFLKMNSLQIDDTAIYYCAKRLGRIFYYAMDYWGQGTSVT 237

Qy    236 VS 237
          ||
Db    238 VS 239
```

```
Search completed: May 13, 1999, 10:02:41
Job time: 1499 sec
```

```
                    GenCore version 4.5
          Copyright (c) 1993 - 1998  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         May 13, 1999, 10:03:48 ; Search time 21.53 Seconds
                                         (without alignments)
                                         417.579 Million cell updates/sec

Title:          US-08-704-178-4
Perfect score:  1271
Sequence:       1 MQLTQSPAIMSASPGEKVTM..........SAYYAMYYWGQGTSVTVSSY 240

Scoring table:  BLOSUM62

Searched:       116738 seqs, 37460341 residues

Database :      PIR_58:*
                1:  pir1:*
                2:  pir2:*
                3:  pir3:*
                4:  pir4:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                SUMMARIES
                  %
Result           Query
  No.   Score    Match Length DB  ID                              Description
---------------------------------------------------------------------------------
    1     560     44.1    287  3  PC4402                          hapten-specific si
    2     526     41.4    137  2  S52445                          Ig heavy chain V r
    3   520.5     41.0    120  2  A34871                          Ig kappa chain V r
    4     496     39.0    107  2  PT0398                          Ig light chain V r
    5     492     38.7    107  2  PT0402                          Ig light chain V r
    6   491.5     38.7    235  2  S25058                          Ig kappa chain - m
    7   487.5     38.4    136  2  S04576                          Ig heavy chain pre
    8     486     38.2    107  2  PT0403                          Ig light chain V r
    9     482     37.9    107  2  PT0400                          Ig light chain V r
   10     482     37.9    107  2  PT0401                          Ig light chain V r
   11     480     37.8     99  2  S51210                          Ig kappa chain V r
   12     480     37.8    107  2  PT0397                          Ig light chain V r
   13     475     37.4    120  2  S03471                          Ig heavy chain V-D
   14     474     37.3    107  2  PT0395                          Ig light chain V r
   15     470     37.0    178  2  S29594                          Ig gamma chain (WM
   16     469     36.9    107  2  A30562                          Ig kappa chain V r
   17     468     36.8    107  2  B30562                          Ig kappa chain V r
   18   464.5     36.5    122  2  S06823                          Ig heavy chain V r
   19     463     36.4    118  2  S25174                          Ig heavy chain V r
   20     461     36.3    130  1  JL0079                          Ig kappa chain pre
   21     460     36.2    107  2  PT0396                          Ig light chain V r
   22     460     36.2    117  2  S17586                          Ig heavy chain V r
   23     459     36.1    107  2  PT0399                          Ig light chain V r
   24   458.5     36.1    268  2  A56446                          Ig heavy chain V r
   25     458     36.0    103  2  S29591                          Ig kappa chain V r
   26   456.5     35.9    118  2  G37267                          Ig heavy chain V r
   27     456     35.9    106  2  G27887                          Ig kappa chain V r
   28     456     35.9    106  2  PS0071                          Ig kappa chain V r
   29     456     35.9    108  2  PL0278                          Ig kappa chain V r
   30   455.5     35.8    108  2  G30560                          Ig kappa chain V r
   31     455     35.8    116  2  S24289                          Ig gamma chain V r
   32     454     35.7    104  2  B49049                          Ig light chain V r
   33     453     35.6    130  2  A32513                          Ig kappa chain pre
   34     452     35.6    108  2  PL0276                          Ig kappa chain V r
   35     452     35.6    108  2  PL0277                          Ig kappa chain V r
   36     450     35.4    106  2  B54378                          Ig light chain V r
   37     445     35.0    130  2  B32456                          Ig kappa chain pre
   38     444     34.9    107  2  PC4405                          anti-AFP antibody
   39     443     34.9    107  2  PT0406                          Ig light chain V r
   40     443     34.9    120  2  S03484                          Ig heavy chain V-D
   41     442     34.8    140  2  PL0013                          Ig kappa chain pre
   42     442     34.8    446  2  S40295                          Ig gamma-2a chain
   43     442     34.8    246  2  S38950                          Ig gamma chain - m
   44     441     34.7    107  2  S11119                          Ig kappa chain V r
   45     440     34.6    107  2  S11118                          Ig kappa chain V r

                                ALIGNMENTS

RESULT   1
PC4402
hapten-specific single-chain antibody variable fragment and alkaline phosphatase fusi
C;Species: Escherichia coli
C;Date: 03-Sep-1998 #sequence_revision 03-Sep-1998 #text_change 03-Sep-1998
C;Accession: PC4402
R;Suzuki, C.; Ueda, H.; Suzuki, E.; Nagamune, T.
J. Biochem. 122, 322-329, 1997
A;Title: Construction, bacterial expression, and characterization of hapten-specific
A;Reference number: PC4402
A;Accession: PC4402
A;Status: preliminary
A;Residues: 1-287 <SUZ>


  Query Match             44.1%;  Score 560;  DB 3;  Length 287;
  Best Local Similarity   47.8%;  Pred. No. 8.6e-36;
  Matches  121;  Conservative   35;  Mismatches   79;  Indels   18;  Gaps    5;

Qy      3 LTQSPAIMSASPGEKVTMTCSASS---SVSN-MHWYQQKSSTSPKLWVYDTSKLASGVPG 58
          :||  | :: |||| ||:|| :|:   : ||  :| |:|        :  |:  | |||
Db     29 VTQESA-LTTSPGETVTLTCRSSTGAVTTSNYANWVQEKPDHLFTGLIGGTNNRAPGVPA 87

Qy     59 RFSGSGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIKGSTSGSGKSSEGK 118
          |||||  |:  :|||:  : || | |:|       : || |||| :  |   : |    |
Db     88 RFSGSLIGDKAALTITGAQTEDEAIYFCALWYSNHWVFGGGTKLTVLSSADDAKKDDAKK 147

Qy    119 G------------VQLQQSGVELVRGGALVKLSCKASDFNIKDYYIHWVKQRPEQGLEWI 166
                       ||||| | |||: || |||||||| :    |::||||||| :|||||
Db    148 DDAKKDDAKKDGQVQLQQPGAELVKPGASVKLSCKASGYTFTSYWMHWVKQRPGRGLEWI 207

Qy    167 GWIHPENGNTVYDPKFQGKASITADTSSNAAYLQLSSLTSEDTAVYYCASYYYYSAYYAM 226
          | | | :| | |: ||: ||::| |  |: ||:|||||||||:|||||| | || : |
Db    208 GRIDPNSGGTKYNEKFKSKATLTVDKPSSTAYMQLSSLTSEDSAVYYCARYDYYGSSY-F 266

Qy    227 YYWGQGTSVTVSS 239
           ||||||::||||
Db    267 DYWGQGTTLTVSS 279


RESULT   2
S52445
Ig heavy chain V region precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 08-May-1995 #sequence_revision 21-Jul-1995 #text_change 08-Sep-1997
C;Accession: S52445
R;Berdoz, J.; Kraehenbuhl, J.P.
submitted to the EMBL Data Library, November 1994
A;Description: Specific amplification by the polymerase chain reaction of rearranged
A;Reference number: S52445
A;Accession: S52445
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-137 <BER>
A;Cross-references: EMBL:X82690; NID:g673439; PID:g673440
C;Genetics:
A;Introns: 16/1
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin


  Query Match             41.4%;  Score 526;  DB 2;  Length 137;
  Best Local Similarity   80.6%;  Pred. No. 1.5e-33;
  Matches  104;  Conservative    5;  Mismatches   14;  Indels    6;  Gaps    2;

Qy    111 SGKSSEGKGVQLQQSGVELVRGGALVKLSCKASDFNIKDYYIHWVKQRPEQGLEWIGWIH 170
          :| :||   ||||||| |||| ||||||||||| |||||||::|||||||||||||||| 
Db     15 TGVNSE---VQLQQSGAELVRPGALVKLSCKASGFNIKDYYMYWVKQRPEQGLEWIGWID 71

Qy    171 PENGNTVYDPKFQGKASITADTSSNAAYLQLSSLTSEDTAVYYCASYYYYSAYYAMYYWG 230
          ||||||||||||||||||||||||| |||||||| ||||||||||   ||   |   |||
Db     72 PENGNTVYDPKFQGKASITADTSSNTAYLQLSSLASEDTAVYYCA---YYGTSYWFPYWG 128

Qy    231 QGTSVTVSS 239
          ||| ||||:
Db    129 QGTLVTVSA 137


RESULT   3
A34871
Ig kappa chain V region - mouse
C;Species: Mus musculus (house mouse)
C;Date: 15-Feb-1991 #sequence_revision 13-Sep-1991 #text_change 20-Mar-1998
C;Accession: A34871
R;Chaudhary, V.K.; Batra, J.K.; Gallo, M.G.; Willingham, M.C.; FitzGerald, D.J.; Pastan,
Proc. Natl. Acad. Sci. U.S.A. 87, 1066-1070, 1990
A;Title: A rapid method of cloning functional variable-region antibody genes in Escheric
A;Reference number: A34871; MUID:90138938
A;Accession: A34871
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-120 <CHA>
A;Cross-references: GB:M34000
A;Note: the authors translated the codon TTT for residue 88 as Pro
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin


  Query Match             41.0%;  Score 520.5;  DB 2;  Length 120;
  Best Local Similarity   86.1%;  Pred. No. 3.3e-33;
  Matches   99;  Conservative    8;  Mismatches    7;  Indels    1;  Gaps    1;

Qy      3 LTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRFSG 62
          |||||||||||||||||||||| |:|| |||||||||||||||:|:||||||||||||||
Db      4 LTQSPAIMSASPGEKVTMTCSAISTVSYMHWYQQKSSTSPKLWIYETSKLASGVPGRFSG 63

Qy     63 SGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIK-GSTSGSGKSSE 116
          |||||||||||||||||| |||||:||||||:||| |||||:| | :||||  |:
Db     64 SGSGNSYSLTISSMEAEDVATYYCFQGSGYPYTFGGGTKLELKEGKSSGSGSESK 118

RESULT   4
PT0398
Ig light chain V region (S107/VH11 group 2-36) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 30-Sep-1993
C;Accession: PT0398
R;Behar, S.M.; Lustgarten, D.L.; Corbet, S.; Scharff, M.D.
J. Exp. Med. 173, 731-741, 1991
A;Title: Characterization of somatically mutated S107 VH11-encoded anti-DNA autoantibodi
A;Reference number: PT0376
A;Accession: PT0398
A;Molecule type: DNA
A;Residues: 1-107 <BEH>
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: immunoglobulin


  Query Match             39.0%;  Score 496;  DB 2;  Length 107;
  Best Local Similarity   90.3%;  Pred. No. 2.1e-31;
  Matches   93;  Conservative    5;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      3 LTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRFSG 62
          ||||||||| ||||||||||||||||| |||||||||||||||:||||||||||||||||
Db      4 LTQSPAIMSTSPGEKVTMTCSASSSVSYMHWYQQKSSTSPKLWIYDTSKLASGVPGRFSG 63

Qy     63 SGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIK 105
          |||||||||||||||||| |||:|: ||||| |||:|||||:|
Db     64 SGSGNSYSLTISSMEAEDVATYFCFXGSGYPLTFGAGTKLELK 106

RESULT   5
PT0402
Ig light chain V region (S107/VH11 group 2-17) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 30-Sep-1993
C;Accession: PT0402
R;Behar, S.M.; Lustgarten, D.L.; Corbet, S.; Scharff, M.D.
J. Exp. Med. 173, 731-741, 1991
A;Title: Characterization of somatically mutated S107 VH11-encoded anti-DNA autoantib
A;Reference number: PT0376
A;Accession: PT0402
A;Molecule type: DNA
A;Residues: 1-107 <BEH>
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: immunoglobulin


  Query Match             38.7%;  Score 492;  DB 2;  Length 107;
  Best Local Similarity   90.3%;  Pred. No. 4.3e-31;
  Matches   93;  Conservative    4;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      3 LTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRFSG 62
          ||||||||||| ||||||||||||||| |||||||||||||||:|||||||||||| |||
Db      4 LTQSPAIMSASXGEKVTMTCSASSSVSYMHWYQQKSSTSPKLWIYDTSKLASGVPGHFSG 63

Qy     63 SGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIK 105
          || ||||||||||||||| |||||:|||||| |||:|||||:|
Db     64 SGXGNSYSLTISSMEAEDVATYYCFQGSGYPLTFGAGTKLELK 106

RESULT   6
S25058
Ig kappa chain - mouse
C;Species: Mus musculus (house mouse)
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 09-Sep-1997
C;Accession: S25058
R;Fischer, R.; Voss, A.; Niersbach, M.; Munziker, W.; Hirsch, H.J.; Kreuzaler, F.
submitted to the EMBL Data Library, July 1992
A;Description: Production of a Tobacco mosaic virus (TMV) inactivating neotop specifi
A;Reference number: S25057
A;Accession: S25058
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-235 <FIS>
A;Cross-references: EMBL:X67211; NID:g54828; PID:g54829


  Query Match             38.7%;  Score 491.5;  DB 2;  Length 235;
  Best Local Similarity   49.8%;  Pred. No. 1.1e-30;
  Matches  119;  Conservative   20;  Mismatches   65;  Indels   35;  Gaps    6;

Qy      3 LTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRFSG 62
          ||||||||||||||||||||||||||| | |||||| |||| |:||||||||||||||||
Db     26 LTQSPAIMSASPGEKVTMTCSASSSVSKMQWYQQKSGTSPKRWIYDTSKLASGVPGRFSG 85

Qy     63 SGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIKGSTSGSGKSSEGKGVQL 122
          |||| ||||||||||||||||||| | |  | |||:|||||:| : :    |        
Db     86 SGSGTSYSLTISSMEAEDAATYYCQQWSSNPLTFGAGTKLELKRADAAPTVSI------F 139

Qy    123 QQSGVELVRGGALVKLSCKASDFNIKDYYIHWVKQRPEQGLEWIGWIHPENGNTVYDPKF 182
            |  :|  ||| |   |  ::|  ||  : |     |:          :||        
Db    140 PPSSEQLTSGGASV--VCFLNNFYPKDINVKWKIDGSER----------QNGVL------ 182

Qy    183 QGKASITADTSSNAAYLQLSSLT-SEDTAVYYCASYYYYSAYYAMYYWGQGTSVTVSSY 240
              | |   | :: |   |:|| ::|        |  :::|         ||  | |:
Db    182 ---NSWTDQDSKDSTYSMSSTLTLTKD-------EYERHNSYTCEATHKTSTSPIVKSF 230
```

```
RESULT   7
S04576
Ig heavy chain precursor V region (MRL-histone 7H) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 13-Feb-1998
C;Accession: S04576
R;Kofler, R.; Noonan, D.J.; Strohal, R.; Balderas, R.S.; Moller, N.P.H.; Dixon, F.J.; Th
Eur. J. Immunol. 17, 91-95, 1987
A;Title: Molecular analysis of the murine lupus-associated anti-self response: involveme
A;Reference number: S04573; MUID:87133856
A;Accession: S04576
A;Molecule type: mRNA
A;Residues: 1-136 <KOF>
A;Cross-references: EMBL:X14624; NID:g52029; PID:g52030
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-136/Product: Ig heavy chain V region (fragment) #status predicted <MAT>


  Query Match             38.4%;  Score 487.5;  DB 2;  Length 136;
  Best Local Similarity   76.6%;  Pred. No. 1.2e-30;
  Matches   98;  Conservative    6;  Mismatches   17;  Indels    7;  Gaps    2;

Qy    112 GKSSEGKGVQLQQSGVELVRGGALVKLSCKASDFNIKDYYIHWVKQRPEQGLEWIGWIHP 171
          | :||   ||||||| |||| || ||||| || ||||| |:||||||||||||||||| |
Db     16 GVNSE---VQLQQSGAELVRPGASVKLSCTASGFNIKDDYMHWVKQRPEQGLEWIGWIDP 72

Qy    172 ENGNTVYDPKFQGKASITADTSSNAAYLQLSSLTSEDTAVYYCASYYYYSAYYAMYYWGQ 231
          |||:| |  ||||||::|||||||  ||||||||||||||||| :|      ||| ||||
Db     73 ENGDTQYASKFQGKATMTADTSSNTTYLQLSSLTSEDTAVYYCTTY----GAYAMDYWGQ 128

Qy    232 GTSVTVSS 239
          ||||||||
Db    129 GTSVTVSS 136


RESULT   8
PT0403
Ig light chain V region (S107/VH11 group 2-19) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 30-Sep-1993
C;Accession: PT0403
R;Behar, S.M.; Lustgarten, D.L.; Corbet, S.; Scharff, M.D.
J. Exp. Med. 173, 731-741, 1991
A;Title: Characterization of somatically mutated S107 VH11-encoded anti-DNA autoantibodi
A;Reference number: PT0376
A;Accession: PT0403
A;Molecule type: DNA
A;Residues: 1-107 <BEH>
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: immunoglobulin


  Query Match             38.2%;  Score 486;  DB 2;  Length 107;
  Best Local Similarity   89.3%;  Pred. No. 1.2e-30;
  Matches   92;  Conservative    5;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      3 LTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRFSG 62
          ||||||||||||||||||||||||| |||||||||||||||: |||||||||||||||
Db      4 LTQSPAIMSASPGEKVTMTCSASSSVSYMHWYQQKSSTSPKLWICDTSKLASGVPGRFSG 63

Qy     63 SGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIK 105
          |||| ||||||||||||| | |||:|||||| |||:|||:|:|
Db     64 SGSGKSYSLTISSMEAEDVAIYYCFQGSGYPLTFGAGTKVELK 106


RESULT   9
PT0400
Ig light chain V region (S107/VH11 group 2-1) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 30-Sep-1993
C;Accession: PT0400
R;Behar, S.M.; Lustgarten, D.L.; Corbet, S.; Scharff, M.D.
J. Exp. Med. 173, 731-741, 1991
A;Title: Characterization of somatically mutated S107 VH11-encoded anti-DNA autoantib
A;Reference number: PT0376
A;Accession: PT0400
A;Molecule type: DNA
A;Residues: 1-107 <BEH>
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: immunoglobulin


  Query Match             37.9%;  Score 482;  DB 2;  Length 107;
  Best Local Similarity   88.3%;  Pred. No. 2.5e-30;
  Matches   91;  Conservative    4;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      3 LTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRFSG 62
          |||||||||| |||||||||||||| |||||||||||||||:|| |||||||||||||
Db      4 LTQSPAIMSAFPGEKVTMTCSASSSVSYMHWYQQKSSTSPKLWIYDXSKLASGVPGRFSG 63

Qy     63 SGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIK 105
          |||||||||||||||||| |||||: |||||  ||:|| ||:|
Db     64 SGSGNSYSLTISSMEAEDVATYYCFXGSGYPLXFGAGTXLELK 106


RESULT  10
PT0401
Ig light chain V region (S107/VH11 group 2-2) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 30-Sep-1993
C;Accession: PT0401
R;Behar, S.M.; Lustgarten, D.L.; Corbet, S.; Scharff, M.D.
J. Exp. Med. 173, 731-741, 1991
A;Title: Characterization of somatically mutated S107 VH11-encoded anti-DNA autoantib
A;Reference number: PT0376
A;Accession: PT0401
A;Molecule type: DNA
A;Residues: 1-107 <BEH>
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: immunoglobulin


  Query Match             37.9%;  Score 482;  DB 2;  Length 107;
  Best Local Similarity   88.3%;  Pred. No. 2.5e-30;
  Matches   91;  Conservative    4;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      3 LTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRFSG 62
          ||||||||||||||||||||||||| |||||||||||||||:|||| |||||||||||
Db      4 LTQSPAIMSASPGEKVTMTCSASSSVSYMHWYQQKSSTSPKLWIYDTSXLASGVPGRFSG 63

Qy     63 SGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIK 105
          |||||||||||| ||||| |||||: |||||  ||:|| ||:|
Db     64 SGSGNSYSLTISRMEAEDGATYYCFXGSGYPLXFGAGTXLELK 106


RESULT  11
S51210
Ig kappa chain V region (AC7) - mouse
C;Species: Mus musculus (house mouse)
C;Date: 01-Aug-1995 #sequence_revision 01-Dec-1995 #text_change 09-May-1997
C;Accession: S51210; S58928
R;Jarrin, A.; Andrieux, A.; Chapel, A.; Buchou, T.; Marguerie, G.
FEBS Lett. 354, 169-172, 1994
A;Title: A synthetic peptide with anti-platelet activity derived from a CDR of an ant
A;Reference number: S51210
A;Accession: S51210
A;Molecule type: mRNA
A;Residues: 9-99 <JAR>
A;Accession: S58928
A;Molecule type: protein
A;Residues: 1-20 <JAR2>
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin
```

```
  Query Match             37.8%;  Score 480;  DB 2;  Length 99;
  Best Local Similarity   93.8%;  Pred. No. 3.2e-30;
  Matches   90;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 LTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRFSG 62
          |||||||||||||||||||||||||||:|||||||||||||||:|||||| |||||||||
Db      4 LTQSPAIMSASPGEKVTMTCSASSSVSSMHWYQQKSSTSPKLWIYDTSKLTSGVPGRFSG 63

Qy     63 SGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGS 98
          |||||||||||||||||| |||||:|||||| ||||
Db     64 SGSGNSYSLTISSMEAEDVATYYCFQGSGYPLTFGS 99


RESULT  12
PT0397
Ig light chain V region (S107/VH11 group 2-10) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 30-Sep-1993
C;Accession: PT0397
R;Behar, S.M.; Lustgarten, D.L.; Corbet, S.; Scharff, M.D.
J. Exp. Med. 173, 731-741, 1991
A;Title: Characterization of somatically mutated S107 VH11-encoded anti-DNA autoantibodi
A;Reference number: PT0376
A;Accession: PT0397
A;Molecule type: DNA
A;Residues: 1-107 <BEH>
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: immunoglobulin


  Query Match             37.8%;  Score 480;  DB 2;  Length 107;
  Best Local Similarity   86.4%;  Pred. No. 3.5e-30;
  Matches   89;  Conservative    7;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      3 LTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRFSG 62
          |||||||||||||:||| |||| | :|  ||||||||||||||:||||||||||||||||
Db      4 LTQSPAIMSASPGKKVTXTCSARSRISYXHWYQQKSSTSPKLWIYDTSKLASGVPGRFSG 63

Qy     63 SGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIK 105
          ||||||||||||:||||| |||||:|||||| |||:|||||:|
Db     64 SGSGNSYSLTISTMEAEDGATYYCFQGSGYPLTFGAGTKLELK 106


RESULT  13
S03471
Ig heavy chain V-D-J region (hybridoma G5 Bb 2.2) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 08-Sep-1997
C;Accession: S03471; S07453
R;Rocca-Serra, J.; Matthes, H.W.; Kaartinen, M.; Milstein, C.; Theze, J.; Fougereau, M.
EMBO J. 2, 867-872, 1983
A;Title: Analysis of antibody diversity: V-D-J mRNA nucleotide sequence of four anti-GAT
hypervariable regions.
A;Reference number: S03471; MUID:84057768
A;Accession: S03471
A;Molecule type: mRNA
A;Residues: 7-120 <ROC1>
A;Cross-references: EMBL:X01820; NID:g51833; PID:e29256; PID:g1333983
A;Note: this sequence was determined from the differentiated gene
R;Rocca-Serra, J.; Mazie, J.C.; Moinier, D.; Leclercq, L.; Somme, G.; Theze, J.; Fougere
J. Immunol. 129, 2554-2558, 1982
A;Title: The limited diversity of the mouse gamma-chains anti-GAT repertoire does not se
A;Reference number: S07453; MUID:83058021
A;Accession: S07453
A;Molecule type: protein
A;Residues: 1-43 <ROC2>
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin


  Query Match             37.4%;  Score 475;  DB 2;  Length 120;
```

```
  Best Local Similarity   80.0%;  Pred. No. 9.6e-30;
  Matches   96;  Conservative    4;  Mismatches   18;  Indels    2;  Gaps    1;

Qy    120 VQLQQSGVELVRGGALVKLSCKASDFNIKDYYIHWVKQRPEQGLEWIGWIHPENGNTVYD 179
          ||||||| |||: || ||||| || ||||| |:||||||||||||||| | | |||| ||
Db      2 VQLQQSGAELVKPGASVKLSCTASGFNIKDTYMHWVKQRPEQGLEWIGRIDPANGNTKYD 61

Qy    180 PKFQGKASITADTSSNAAYLQLSSLTSEDTAVYYCASYYYYSAYYAMYYWGQGTSVTVSS 239
          |||||||:|||||||| |||||||||||||||||||  :      || ||||||||||||
Db     62 PKFQGKATITADTSSNTAYLQLSSLTSEDTAVYYCARGWLRRD--AMDYWGQGTSVTVSS 119


RESULT  14
PT0395
Ig light chain V region (S107/VH11 group 2-16) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 30-Sep-1993
C;Accession: PT0395
R;Behar, S.M.; Lustgarten, D.L.; Corbet, S.; Scharff, M.D.
J. Exp. Med. 173, 731-741, 1991
A;Title: Characterization of somatically mutated S107 VH11-encoded anti-DNA autoantib
A;Reference number: PT0376
A;Accession: PT0395
A;Molecule type: DNA
A;Residues: 1-107 <BEH>
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: immunoglobulin


  Query Match             37.3%;  Score 474;  DB 2;  Length 107;
  Best Local Similarity   85.4%;  Pred. No. 1e-29;
  Matches   88;  Conservative    7;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      3 LTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRFSG 62
          | ||||||||||||||||||||||||| |||||||||||||:|:||||||||||||||||
Db      4 LIQSPAIMSASPGEKVTMTCSASSSVSYMHWYQQKSSTSPKVWIYDTSKLASGVPGRFSG 63

Qy     63 SGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIK 105
          |||||||:||||:||||| |||||: | |:| ||| || ||:|
Db     64 SGSGNSYALTISNMEAEDVATYYCFXGXGFPLTFGGGTTLELK 106


RESULT  15
S29594
Ig gamma chain (WM65) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 10-Sep-1997
C;Accession: S29594
R;Seymour, R.
submitted to the EMBL Data Library, February 1991
A;Reference number: S29593
A;Accession: S29594
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-178 <SEY>
A;Cross-references: EMBL:X57857; NID:g52590; PID:g52591
C;Keywords: immunoglobulin


  Query Match             37.0%;  Score 470;  DB 2;  Length 178;
  Best Local Similarity   74.4%;  Pred. No. 3.5e-29;
  Matches   96;  Conservative    7;  Mismatches   20;  Indels    6;  Gaps    2;

Qy    111 SGKSSEGKGVQLQQSGVELVRGGALVKLSCKASDFNIKDYYIHWVKQRPEQGLEWIGWIH 170
          :| :||   ||||||| |||: || ||||| || ||||| |:|||||||:||||||| | 
Db      9 TGVNSE---VQLQQSGAELVKPGASVKLSCTASGFNIKDTYMHWVKQRPKQGLEWIGRID 65

Qy    171 PENGNTVYDPKFQGKASITADTSSNAAYLQLSSLTSEDTAVYYCASYYYYSAYYAMYYWG 230
          | || | |||||||||:||||||:| ||||||||||||||||||    |    | | |||
Db     66 PANGYTEYDPKFQGKATITADTSTNTAYLQLSSLTSEDTAVYYCTGGNY---AYGMDYWG 122

Qy    231 QGTSVTVSS 239
          |||||||||
Db    123 QGTSVTVSS 131


Search completed: May 13, 1999, 10:03:48
Job time: 1380 sec
```

```
                    GenCore version 4.5
          Copyright (c) 1993 - 1998  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         May 13, 1999, 10:05:48 ; Search time 16.53 Seconds
                                   (without alignments)
                                   389.696 Million cell updates/sec

Title:          US-08-704-178-4
Perfect score:  1271
Sequence:       1 MQLTQSPAIMSASPGEKVTM..........SAYYAMYYWGQGTSVTVSSY 240

Scoring table:  BLOSUM62

Searched:       74019 seqs, 26840295 residues

Database :      SwissProt_36:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                 SUMMARIES
                        %
Result                Query
  No.     Score   Match Length DB  ID                     Description
-------------------------------------------------------------------------------
    1      450    35.4    107  1  KV6F_MOUSE             P04940 mus musculu
    2      446    35.1    107  1  KV6J_MOUSE             P04944 mus musculu
    3      445    35.0    107  1  KV6I_MOUSE             P04943 mus musculu
    4      444    34.9    107  1  KV6G_MOUSE             P04941 mus musculu
    5      444    34.9    107  1  KV6H_MOUSE             P04942 mus musculu
    6      412    32.4    107  1  KV6B_MOUSE             P01676 mus musculu
    7      410    32.3    108  1  KV6K_MOUSE             P04945 mus musculu
    8      409    32.2    107  1  KV6A_MOUSE             P01675 mus musculu
    9      408    32.1    107  1  KV6D_MOUSE             P01678 mus musculu
   10      407    32.0    107  1  KV6C_MOUSE             P01677 mus musculu
   11      406    31.9    120  1  HV03_MOUSE             P01747 mus musculu
   12    401.5    31.6    140  1  HV02_MOUSE             P01746 mus musculu
   13      401    31.5    107  1  KV6E_MOUSE             P01679 mus musculu
   14      400    31.5    139  1  HV07_MOUSE             P01751 mus musculu
   15      395    31.1    129  1  KV4A_MOUSE             P01680 mus musculu
   16      370    29.1    117  1  HV13_MOUSE             P01757 mus musculu
   17    367.5    28.9    118  1  HV51_MOUSE             P06330 mus musculu
   18      366    28.8    121  1  HV01_MOUSE             P01745 mus musculu
   19      365    28.7    117  1  HV12_MOUSE             P01756 mus musculu
   20      363    28.6    137  1  HV11_MOUSE             P01755 mus musculu
   21    361.5    28.4    120  1  HV50_MOUSE             P06329 mus musculu
   22      361    28.4    138  1  HV48_MOUSE             P03980 mus musculu
   23      356    28.0    117  1  HV52_MOUSE             P06327 mus musculu
   24    352.5    27.7    143  1  HV1C_HUMAN             P01744 homo sapien
   25      346    27.2    117  1  HV06_MOUSE             P01750 mus musculu
   26      340    26.8    136  1  HV15_MOUSE             P01759 mus musculu
   27    337.5    26.6    117  1  HV09_MOUSE             P01753 mus musculu
   28      336    26.4    117  1  HV04_MOUSE             P01748 mus musculu
   29      330    26.0    117  1  HV14_MOUSE             P01758 mus musculu
   30    328.5    25.8    108  1  KV1R_HUMAN             P01610 homo sapien
   31    326.5    25.7    108  1  KV5K_MOUSE             P01644 mus musculu
   32      326    25.6    117  1  HV49_MOUSE             P06328 mus musculu
   33    325.5    25.6    108  1  KV1B_HUMAN             P01594 homo sapien
   34    325.5    25.6    108  1  KV1V_HUMAN             P04430 homo sapien
   35      325    25.6    117  1  HV05_MOUSE             P01749 mus musculu
   36    324.5    25.5    108  1  KV5O_MOUSE             P01648 mus musculu
   37    323.5    25.5    108  1  KV5L_MOUSE             P01645 mus musculu
   38    323.5    25.5    108  1  KV5M_MOUSE             P01646 mus musculu
   39    322.5    25.4    108  1  KV1P_HUMAN             P01608 homo sapien
   40    322.5    25.4    129  1  KV1W_HUMAN             P04431 homo sapien
   41      322    25.3    109  1  KV3F_HUMAN             P01624 homo sapien
   42    321.5    25.3    108  1  KV5D_MOUSE             P01636 mus musculu
   43      321    25.3    117  1  HV10_MOUSE             P01754 mus musculu
   44    320.5    25.2    108  1  KV1L_HUMAN             P01604 homo sapien
   45    320.5    25.2    108  1  KV5N_MOUSE             P01647 mus musculu
```

```
                               ALIGNMENTS


RESULT   1
KV6F_MOUSE
ID   KV6F_MOUSE     STANDARD;      PRT;   107 AA.
AC   P04940;
DT   13-AUG-1987 (REL. 05, CREATED)
DT   13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT   01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
DE   IG KAPPA CHAIN V-VI REGION (NQ2-17.4.1).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 83271467.
RA   KAARTINEN M., GRIFFITHS G.M., MARKHAM A.F., MILSTEIN C.;
RL   NATURE 304:320-324(1983).
CC   -!- FUNCTION: ANTI-2-PHENYL OXAZOLONE (PHOX) ANTIBODY.
DR   EMBL; K00735; G196435; -.
DR   HSSP; P01679; 1FVB.
KW   IMMUNOGLOBULIN V REGION; HYBRIDOMA.
FT   DOMAIN        1     23       FRAMEWORK 1.
FT   DOMAIN       24     33       COMPLEMENTARITY-DETERMINING 1.
FT   DOMAIN       34     48       FRAMEWORK 2.
FT   DOMAIN       49     55       COMPLEMENTARITY-DETERMINING 2.
FT   DOMAIN       56     87       FRAMEWORK 3.
FT   DOMAIN       88     96       COMPLEMENTARITY-DETERMINING 3.
FT   DOMAIN       97    106       FRAMEWORK 4.
FT   DISULFID     23     87       BY SIMILARITY.
FT   NON_TER     107    107
SQ   SEQUENCE   107 AA;  11561 MW;  1BC0C1E3 CRC32;


  Query Match             35.4%;  Score 450;  DB 1;  Length 107;
  Best Local Similarity   83.5%;  Pred. No. 2.1e-28;
  Matches   86;  Conservative    6;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      3 LTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRFSG 62
          ||||||||||||:||||||||||||| |||||||| |||| |:||||||||||| ||||
Db      4 LTQSPAIMSASPGQKVTMTCSASSSVSYMHWYQQKSGTSPKRWIYDTSKLASGVPARFSG 63

Qy     63 SGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIK 105
          |||  ||||||:||:||||||||| | |  | |||:|||||:|  .
Db     64 SGSATSYSLTITSMQAEDAATYYCQQWSSNPLTFGAGTKLELK 106


RESULT   2
KV6J_MOUSE
ID   KV6J_MOUSE     STANDARD;      PRT;   107 AA.
AC   P04944;
DT   13-AUG-1987 (REL. 05, CREATED)
DT   13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT   01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
DE   IG KAPPA CHAIN V-VI REGION (NQ5-78.2.6).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 83271467.
RA   KAARTINEN M., GRIFFITHS G.M., MARKHAM A.F., MILSTEIN C.;
RL   NATURE 304:320-324(1983).
CC   -!- FUNCTION: ANTI-2-PHENYL OXAZOLONE (PHOX) ANTIBODY.
DR   EMBL; K00744; G196453; -.
DR   HSSP; P01679; 1FVB.
KW   IMMUNOGLOBULIN V REGION; HYBRIDOMA.
FT   DOMAIN        1     23       FRAMEWORK 1.
FT   DOMAIN       24     33       COMPLEMENTARITY-DETERMINING 1.
FT   DOMAIN       34     48       FRAMEWORK 2.
FT   DOMAIN       49     55       COMPLEMENTARITY-DETERMINING 2.
FT   DOMAIN       56     87       FRAMEWORK 3.
FT   DOMAIN       88     96       COMPLEMENTARITY-DETERMINING 3.
FT   DOMAIN       97    106       FRAMEWORK 4.
FT   DISULFID     23     87       BY SIMILARITY.
FT   NON_TER     107    107
SQ   SEQUENCE   107 AA;  11611 MW;  E926BBEB CRC32;

  Query Match             35.1%;  Score 446;  DB 1;  Length 107;
  Best Local Similarity   83.5%;  Pred. No. 4.3e-28;
  Matches   86;  Conservative    4;  Mismatches   13;  Indels    0;  Gaps    0;

Qy       3 LTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRFSG 62
           |||||||||||||:||||||||||||| |||||||| |||| |:||||||||||| || |
Db       4 LTQSPAIMSASPGQKVTMTCSASSSVSYMHWYQQKSGTSPKRWIYDTSKLASGVPARFXG 63

Qy      63 SGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIK 105
           |||  ||||||:||:||||||||| | |  | ||||||||| |
Db      64 SGSATSYSLTITSMQAEDAATYYCQQWSSNPLTFGSGTKLEXK 106


RESULT   3
KV6I_MOUSE
ID   KV6I_MOUSE     STANDARD;      PRT;   107 AA.
AC   P04943;
DT   13-AUG-1987 (REL. 05, CREATED)
DT   13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT   01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
DE   IG KAPPA CHAIN V-VI REGION (NQ6-8.3.1).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 83271467.
RA   KAARTINEN M., GRIFFITHS G.M., MARKHAM A.F., MILSTEIN C.;
RL   NATURE 304:320-324(1983).
CC   -!- FUNCTION: ANTI-2-PHENYL OXAZOLONE (PHOX) ANTIBODY.
DR   EMBL; K00740; G196445; -.
DR   HSSP; P01679; 1FVB.
KW   IMMUNOGLOBULIN V REGION; HYBRIDOMA.
FT   DOMAIN        1     23       FRAMEWORK 1.
FT   DOMAIN       24     33       COMPLEMENTARITY-DETERMINING 1.
FT   DOMAIN       34     48       FRAMEWORK 2.
FT   DOMAIN       49     55       COMPLEMENTARITY-DETERMINING 2.
FT   DOMAIN       56     87       FRAMEWORK 3.
FT   DOMAIN       88     96       COMPLEMENTARITY-DETERMINING 3.
FT   DOMAIN       97    106       FRAMEWORK 4.
FT   DISULFID     23     87       BY SIMILARITY.
FT   NON_TER     107    107
SQ   SEQUENCE   107 AA;  11572 MW;  2D976197 CRC32;

  Query Match             35.0%;  Score 445;  DB 1;  Length 107;
  Best Local Similarity   82.5%;  Pred. No. 5.1e-28;
  Matches   85;  Conservative    6;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       3 LTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRFSG 62
           |||||||||||||:||||||||||||| |||||||| |||| |:||||||||| | ||||
Db       4 LTQSPAIMSASPGQKVTMTCSASSSVSYMHWYQQKSGTSPKRWIYDTSKLASGXPARFSG 63

Qy      63 SGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIK 105
           |||  ||||||:||:||||||||| | |  | |||:|||||:|
Db      64 SGSATSYSLTITSMQAEDAATYYCQQWSSNPLTFGAGTKLELK 106


RESULT   4
KV6G_MOUSE
ID   KV6G_MOUSE     STANDARD;      PRT;   107 AA.
AC   P04941;
```

```
DT   13-AUG-1987 (REL. 05, CREATED)
DT   13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT   01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
DE   IG KAPPA CHAIN V-VI REGION (NQ2-48.2.2).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 83271467.
RA   KAARTINEN M., GRIFFITHS G.M., MARKHAM A.F., MILSTEIN C.;
RL   NATURE 304:320-324(1983).
CC   -!- FUNCTION: ANTI-2-PHENYL OXAZOLONE (PHOX) ANTIBODY.
DR   EMBL; K00737; G196439; -.
DR   HSSP; P01679; 1FVB.
KW   IMMUNOGLOBULIN V REGION; HYBRIDOMA.
FT   DOMAIN        1     23       FRAMEWORK 1.
FT   DOMAIN       24     33       COMPLEMENTARITY-DETERMINING 1.
FT   DOMAIN       34     48       FRAMEWORK 2.
FT   DOMAIN       49     55       COMPLEMENTARITY-DETERMINING 2.
FT   DOMAIN       56     87       FRAMEWORK 3.
FT   DOMAIN       88     96       COMPLEMENTARITY-DETERMINING 3.
FT   DOMAIN       97    106       FRAMEWORK 4.
FT   DISULFID     23     87       BY SIMILARITY.
FT   NON_TER     107    107
SQ   SEQUENCE   107 AA;  11556 MW;  1AB8295D CRC32;

  Query Match             34.9%;  Score 444;  DB 1;  Length 107;
  Best Local Similarity   82.5%;  Pred. No. 6.1e-28;
  Matches   85;  Conservative    6;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       3 LTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRFSG 62
           |||||||||||||:||||||||||||| |||||||| |||| |:||||||||||| ||||
Db       4 LTQSPAIMSASPGQKVTMTCSASSSVSYMHWYQQKSGTSPKRWIYDTSKLASGVPARFSG 63

Qy      63 SGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIK 105
           |||  ||||||:||:||||||||| | |  | |||:|||| :|
Db      64 SGSATSYSLTITSMQAEDAATYYCQQWSSNPLTFGAGTKLXLK 106


RESULT   5
KV6H_MOUSE
ID   KV6H_MOUSE     STANDARD;      PRT;   107 AA.
AC   P04942;
DT   13-AUG-1987 (REL. 05, CREATED)
DT   13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT   01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
DE   IG KAPPA CHAIN V-VI REGION (NQ5-61.1.2).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 83271467.
RA   KAARTINEN M., GRIFFITHS G.M., MARKHAM A.F., MILSTEIN C.;
RL   NATURE 304:320-324(1983).
CC   -!- FUNCTION: ANTI-2-PHENYL OXAZOLONE (PHOX) ANTIBODY.
DR   EMBL; K00739; G196443; -.
DR   HSSP; P01679; 1FVB.
KW   IMMUNOGLOBULIN V REGION; HYBRIDOMA.
FT   DOMAIN        1     23       FRAMEWORK 1.
FT   DOMAIN       24     33       COMPLEMENTARITY-DETERMINING 1.
FT   DOMAIN       34     48       FRAMEWORK 2.
FT   DOMAIN       49     55       COMPLEMENTARITY-DETERMINING 2.
FT   DOMAIN       56     87       FRAMEWORK 3.
FT   DOMAIN       88     96       COMPLEMENTARITY-DETERMINING 3.
FT   DOMAIN       97    106       FRAMEWORK 4.
FT   DISULFID     23     87       BY SIMILARITY.
FT   NON_TER     107    107
SQ   SEQUENCE   107 AA;  11605 MW;  1CF3CE8D CRC32;
```

```
  Query Match             34.9%;  Score 444;  DB 1;  Length 107;
  Best Local Similarity   82.5%;  Pred. No. 6.1e-28;
  Matches   85;  Conservative    6;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        3 LTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRFSG 62
            |||||||||||||:||||||||||||| |||||||| |||| |:|||||| |||| ||||
Db        4 LTQSPAIMSASPGQKVTMTCSASSSVSYMHWYQQKSGTSPKRWIYDTSKLDSGVPARFSG 63

Qy       63 SGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIK 105
            |||  ||||||:||:||||||||| | |  | |||:|||||:|
Db       64 SGSATSYSLTITSMQAEDAATYYCQQWSSNPLTFGAGTKLELK 106


RESULT   6
KV6B_MOUSE
ID   KV6B_MOUSE     STANDARD;      PRT;   107 AA.
AC   P01676;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
DE   IG KAPPA CHAIN V-VI REGION (XRPC 24).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 79082830.
RA   RAO D.N., RUDIKOFF S., POTTER M.;
RL   BIOCHEMISTRY 17:5555-5559(1978).
CC   -!- THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT BIND GALACTAN.
DR   PIR; A01941; KVMSX4.
DR   HSSP; P01679; 2FBJ.
KW   IMMUNOGLOBULIN V REGION.
FT   DOMAIN        1     23       FRAMEWORK 1.
FT   DOMAIN       24     33       COMPLEMENTARITY-DETERMINING 1.
FT   DOMAIN       34     48       FRAMEWORK 2.
FT   DOMAIN       49     55       COMPLEMENTARITY-DETERMINING 2.
FT   DOMAIN       56     87       FRAMEWORK 3.
FT   DOMAIN       88     96       COMPLEMENTARITY-DETERMINING 3.
FT   DOMAIN       97    106       FRAMEWORK 4.
FT   DISULFID     23     87       BY SIMILARITY.
FT   NON_TER     107    107
SQ   SEQUENCE   107 AA;  11584 MW;  91B2CD6E CRC32;

  Query Match             32.4%;  Score 412;  DB 1;  Length 107;
  Best Local Similarity   80.8%;  Pred. No. 1.7e-25;
  Matches   84;  Conservative    6;  Mismatches   12;  Indels    2;  Gaps    2;

Qy        3 LTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRFSG 62
            ||||||| :|| |:|||:||||||||| |||||||| |||| |:|: |||||||| ||||
Db        4 LTQSPAITAASLGQKVTITCSASSSVSYMHWYQQKSGTSPKPWIYEISKLASGVPARFSG 63

Qy       63 SGSGNSYSLTISSMEAEDAATYYCYQGSGYP-FTFGSGTKLEIK 105
            |||| ||||||||||||||| ||| | : ||  |||||||||||
Db       64 SGSGTSYSLTISSMEAEDAAIYYCQQWN-YPLITFGSGTKLEIK 106

RESULT   7
KV6K_MOUSE
ID   KV6K_MOUSE     STANDARD;      PRT;   108 AA.
AC   P04945;
DT   13-AUG-1987 (REL. 05, CREATED)
DT   13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT   01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
DE   IG KAPPA CHAIN V-VI REGION (NQ2-6.1).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 83271467.
RA   KAARTINEN M., GRIFFITHS G.M., MARKHAM A.F., MILSTEIN C.;
RL   NATURE 304:320-324(1983).
DR   EMBL; K00746; G554059; -.
DR   HSSP; P01679; 1BAF.
KW   IMMUNOGLOBULIN V REGION; HYBRIDOMA.
FT   DOMAIN        1     23       FRAMEWORK 1.
FT   DOMAIN       24     33       COMPLEMENTARITY-DETERMINING 1.
FT   DOMAIN       34     48       FRAMEWORK 2.
FT   DOMAIN       49     55       COMPLEMENTARITY-DETERMINING 2.
FT   DOMAIN       56     87       FRAMEWORK 3.
FT   DOMAIN       88     98       COMPLEMENTARITY-DETERMINING 3.
FT   DOMAIN       99    108       FRAMEWORK 4.
FT   DISULFID     23     87       BY SIMILARITY.
FT   NON_TER     108    108
SQ   SEQUENCE   108 AA;  11713 MW;  8B48A204 CRC32;

  Query Match             32.3%;  Score 410;  DB 1;  Length 108;
  Best Local Similarity   77.1%;  Pred. No. 2.5e-25;
  Matches   81;  Conservative    9;  Mismatches   13;  Indels    2;  Gaps    1;

Qy        3 LTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRFSG 62
            |||||||||||||:||||||||||||| |:|||||  :||:| :|||| |||||| ||||
Db        4 LTQSPAIMSASPGQKVTMTCSASSSVSYMYWYQQKPGSSPRLLIYDTSNLASGVPVRFSG 63

Qy       63 SGSGNSYSLTISSMEAEDAATYYCYQGSGYP--FTFGSGTKLEIK 105
            |||  ||||||: |:||||||||| | | ||    |||:|||||:|
Db       64 SGSATSYSLTITRMQAEDAATYYCQQWSSYPPMLTFGAGTKLELK 108


RESULT   8
KV6A_MOUSE
ID   KV6A_MOUSE     STANDARD;      PRT;   107 AA.
AC   P01675;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
DE   IG KAPPA CHAIN V-VI REGION (XRPC 44).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 79082830.
RA   RAO D.N., RUDIKOFF S., POTTER M.;
RL   BIOCHEMISTRY 17:5555-5559(1978).
CC   -!- THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT BIND GALACTAN.
DR   PIR; A01941; KVMSX4.
DR   HSSP; P01679; 2FBJ.
KW   IMMUNOGLOBULIN V REGION.
FT   DOMAIN        1     23       FRAMEWORK 1.
FT   DOMAIN       24     33       COMPLEMENTARITY-DETERMINING 1.
FT   DOMAIN       34     48       FRAMEWORK 2.
FT   DOMAIN       49     55       COMPLEMENTARITY-DETERMINING 2.
FT   DOMAIN       56     87       FRAMEWORK 3.
FT   DOMAIN       88     96       COMPLEMENTARITY-DETERMINING 3.
FT   DOMAIN       97    106       FRAMEWORK 4.
FT   DISULFID     23     87       BY SIMILARITY.
FT   NON_TER     107    107
SQ   SEQUENCE   107 AA;  11627 MW;  2E7F7479 CRC32;

  Query Match             32.2%;  Score 409;  DB 1;  Length 107;
  Best Local Similarity   79.8%;  Pred. No. 2.9e-25;
  Matches   83;  Conservative    7;  Mismatches   12;  Indels    2;  Gaps    2;

Qy        3 LTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRFSG 62
            ||||||| :|| |:|||:||||||||| |||||||| |||| |:|: |||||||| ||||
Db        4 LTQSPAITAASLGQKVTITCSASSSVSYMHWYQQKSGTSPKPWIYEISKLASGVPARFSG 63

Qy       63 SGSGNSYSLTISSMEAEDAATYYCYQGSGYP-FTFGSGTKLEIK 105
            |||| ||||||||||||||| ||| | : || :||| |||||||
Db       64 SGSGTSYSLTISSMEAEDAAIYYCQQWN-YPLWTFGGGTKLEIK 106
```

```
RESULT   9
KV6D_MOUSE
ID   KV6D_MOUSE     STANDARD;      PRT;   107 AA.
AC   P01678;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
DE   IG KAPPA CHAIN V-VI REGION (SAPC 10).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 81054757.
RA   RUDIKOFF S., RAO D.N., GLAUDEMANS C.P.J., POTTER M.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 77:4270-4274(1980).
CC   -!- THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT BIND GALACTAN.
DR   PIR; A01941; KVMSX4.
DR   HSSP; P01679; 2FBJ.
KW   IMMUNOGLOBULIN V REGION.
FT   DOMAIN        1     23       FRAMEWORK 1.
FT   DOMAIN       24     33       COMPLEMENTARITY-DETERMINING 1.
FT   DOMAIN       34     48       FRAMEWORK 2.
FT   DOMAIN       49     55       COMPLEMENTARITY-DETERMINING 2.
FT   DOMAIN       56     87       FRAMEWORK 3.
FT   DOMAIN       88     96       COMPLEMENTARITY-DETERMINING 3.
FT   DOMAIN       97    106       FRAMEWORK 4.
FT   DISULFID     23     87       BY SIMILARITY.
FT   NON_TER     107    107
SQ   SEQUENCE   107 AA;  11554 MW;  609EDE13 CRC32;

  Query Match             32.1%;  Score 408;  DB 1;  Length 107;
  Best Local Similarity   79.8%;  Pred. No. 3.5e-25;
  Matches   83;  Conservative    6;  Mismatches   13;  Indels    2;  Gaps    2;

Qy      3 LTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRFSG 62
          ||||||| :|| |:|||:||||||||| |||||||| |||| |:|: |||||||| ||||
Db      4 LTQSPAITAASLGQKVTITCSASSSVSYMHWYQQKSGTSPKPWIYEISKLASGVPARFSG 63

Qy     63 SGSGNSYSLTISSMEAEDAATYYCYQGSGYP-FTFGSGTKLEIK 105
          |||| ||||||||||||||| ||| | : ||  ||| |||||||
Db     64 SGSGTSYSLTISSMEAEDAAIYYCQQWN-YPLITFGGGTKLEIK 106

RESULT  10
KV6C_MOUSE
ID   KV6C_MOUSE     STANDARD;      PRT;   107 AA.
AC   P01677;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
DE   IG KAPPA CHAIN V-VI REGION (TEPC 601 AND TEPC 191).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE (TEPC 601).
RX   MEDLINE; 79082830.
RA   RAO D.N., RUDIKOFF S., POTTER M.;
RL   BIOCHEMISTRY 17:5555-5559(1978).
RN   [2]
RP   SEQUENCE (TEPC 191).
RX   MEDLINE; 81054757.
RA   RUDIKOFF S., RAO D.N., GLAUDEMANS C.P.J., POTTER M.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 77:4270-4274(1980).
CC   -!- THE TWO SEQUENCES ARE IDENTICAL.
CC   -!- THESE CHAINS WERE ISOLATED FROM MYELOMA PROTEINS THAT BIND
CC       GALACTAN.
DR   PIR; A01941; KVMSX4.
DR   HSSP; P01679; 2FBJ.
KW   IMMUNOGLOBULIN V REGION.
FT   DOMAIN        1     23       FRAMEWORK 1.
FT   DOMAIN       24     33       COMPLEMENTARITY-DETERMINING 1.
FT   DOMAIN       34     48       FRAMEWORK 2.
FT   DOMAIN       49     55       COMPLEMENTARITY-DETERMINING 2.
FT   DOMAIN       56     87       FRAMEWORK 3.
FT   DOMAIN       88     96       COMPLEMENTARITY-DETERMINING 3.
FT   DOMAIN       97    106       FRAMEWORK 4.
FT   DISULFID     23     87       BY SIMILARITY.
FT   NON_TER     107    107
SQ   SEQUENCE   107 AA;  11568 MW;  154F6572 CRC32;

  Query Match             32.0%;  Score 407;  DB 1;  Length 107;
  Best Local Similarity   78.8%;  Pred. No. 4.2e-25;
  Matches   82;  Conservative    8;  Mismatches   12;  Indels    2;  Gaps    2;

Qy      3 LTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRFSG 62
          ||||||| :|| |:|||:||||||||| |||||||| |||| |:|: |||||||| ||||
Db      4 LTQSPAITAASLGQKVTITCSASSSVSYMHWYQQKSGTSPKPWIYEISKLASGVPARFSG 63

Qy     63 SGSGNSYSLTISSMEAEDAATYYCYQGSGYP-FTFGSGTKLEIK 105
          |||| ||||||||||||||| ||| | : ||  |||:|||||:|
Db     64 SGSGTSYSLTISSMEAEDAAIYYCQQWN-YPLITFGAGTKLELK 106

RESULT  11
HV03_MOUSE
ID   HV03_MOUSE     STANDARD;      PRT;   120 AA.
AC   P01747;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   21-JUL-1986 (REL. 01, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN V REGION (36-65).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 83131846.
RA   SIEKEVITZ M., GEFTER M.L., BRODEUR P., RIBLET R.,
RA   MARSHAK-ROTHSTEIN A.;
RL   EUR. J. IMMUNOL. 12:1023-1032(1982).
CC   -!- FROM ANALYSIS OF THE SIZES OF SEVERAL OTHER DIFFERENTIATED GENES
CC       THAT HYBRIDIZE TO THIS ONE, THE AUTHORS CONCLUDE THAT ALL OF
CC       THESE V REGIONS HAVE REARRANGED TO THE SAME J SEGMENT, JH2.
DR   PIR; A02028; HVMSG7.
DR   HSSP; P01789; 6FAB.
KW   IMMUNOGLOBULIN V REGION; ANTIARSONATE ANTIBODY; HYBRIDOMA.
FT   NON_TER     120    120
SQ   SEQUENCE   120 AA;  13307 MW;  BBA8CCA1 CRC32;

  Query Match             31.9%;  Score 406;  DB 1;  Length 120;
  Best Local Similarity   63.3%;  Pred. No. 5.7e-25;
  Matches   76;  Conservative   16;  Mismatches   28;  Indels    0;  Gaps    0;

Qy    120 VQLQQSGVELVRGGALVKLSCKASDFNIKDYYIHWVKQRPEQGLEWIGWIHPENGNTVYD 179
          ||||||| |||| |: ||:||||| :    | |:|||||| |||||||:|:| || | |:
Db      1 VQLQQSGAELVRAGSSVKMSCKASGYTFTSYGINWVKQRPGQGLEWIGYINPGNGYTKYN 60

Qy    180 PKFQGKASITADTSSNAAYLQLSSLTSEDTAVYYCASYYYYSAYYAMYYWGQGTSVTVSS 239
           ||:|| ::| | ||: ||:|| ||||||:|||:||   ||   |   ||||||::||||
Db     61 EKFKGKTTLTVDKSSSTAYMQLRSLTSEDSAVYFCARSVYYGGSYYFDYWGQGTTLTVSS 120

RESULT  12
HV02_MOUSE
ID   HV02_MOUSE     STANDARD;      PRT;   140 AA.
AC   P01746;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN PRECURSOR V REGION (93G7).
```

```
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/J;
RX   MEDLINE; 82152818.
RA   SIMS J., RABBITTS T.H., ESTESS P., SLAUGHTER C., TUCKER P.W.,
RA   CAPRA J.D.;
RL   SCIENCE 216:309-311(1982).
DR   EMBL; J00493; G195007; -.
DR   PIR; A02028; HVMSG7.
DR   HSSP; P01789; 6FAB.
KW   IMMUNOGLOBULIN V REGION; ANTIARSONATE ANTIBODY; HYBRIDOMA; SIGNAL.
FT   SIGNAL        1     19
FT   CHAIN        20    140       IG HEAVY CHAIN V REGION (93G7).
FT   NON_TER     140    140
SQ   SEQUENCE   140 AA;  15514 MW;  0700D5C8 CRC32;

  Query Match             31.6%;  Score 401.5;  DB 1;  Length 140;
  Best Local Similarity   60.5%;  Pred. No. 1.5e-24;
  Matches   78;  Conservative   17;  Mismatches   31;  Indels    3;  Gaps    1;

Qy    111 SGKSSEGKGVQLQQSGVELVRGGALVKLSCKASDFNIKDYYIHWVKQRPEQGLEWIGWIH 170
          :|  ||   ||||||| |||| |: ||:||||| :    | |:|||||| |||||||:|:
Db     15 AGVHSE---VQLQQSGAELVRAGSSVKMSCKASGYTFTSYGINWVKQRPGQGLEWIGYIN 71

Qy    171 PENGNTVYDPKFQGKASITADTSSNAAYLQLSSLTSEDTAVYYCASYYYYSAYYAMYYWG 230
          | ||   |: ||:|| ::| | ||: ||:|| ||||||:|||:||  :||   |   |||
Db     72 PGNGYINYNEKFKGKTTLTVDKSSSTAYMQLRSLTSEDSAVYFCARSHYYGGSYDFDYWG 131

Qy    231 QGTSVTVSS 239
          ||| :||||
Db    132 QGTPLTVSS 140


RESULT  13
KV6E_MOUSE
ID   KV6E_MOUSE     STANDARD;      PRT;   107 AA.
AC   P01679;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
DE   IG KAPPA CHAIN V-VI REGION (J539).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 81054757.
RA   RUDIKOFF S., RAO D.N., GLAUDEMANS C.P.J., POTTER M.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 77:4270-4274(1980).
RN   [2]
RP   X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
RX   MEDLINE; 88217852.
RA   SUH S.W., BHAT T.N., NAVIA M.A., COHEN G.H., RAO D.N., RUDIKOFF S.,
RA   DAVIES D.R.;
RL   PROTEINS 1:74-80(1986).
CC   -!- THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT BIND GALACTAN.
DR   PIR; A01942; KVMSJ5.
DR   PDB; 2FBJ; 15-OCT-90.
KW   IMMUNOGLOBULIN V REGION; 3D-STRUCTURE.
FT   DOMAIN        1     23       FRAMEWORK 1.
FT   DOMAIN       24     33       COMPLEMENTARITY-DETERMINING 1.
FT   DOMAIN       34     48       FRAMEWORK 2.
FT   DOMAIN       49     55       COMPLEMENTARITY-DETERMINING 2.
FT   DOMAIN       56     87       FRAMEWORK 3.
FT   DOMAIN       88     96       COMPLEMENTARITY-DETERMINING 3.
FT   DOMAIN       97    106       FRAMEWORK 4.
FT   DISULFID     23     87
FT   NON_TER     107    107
FT   STRAND        4      7
```

```
FT   STRAND       10     14
FT   TURN         15     16
FT   STRAND       19     25
FT   STRAND       31     37
FT   TURN         39     40
FT   STRAND       44     48
FT   TURN         49     51
FT   STRAND       52     53
FT   TURN         55     56
FT   TURN         59     60
FT   STRAND       61     66
FT   TURN         67     68
FT   STRAND       69     74
FT   HELIX        79     81
FT   STRAND       83     91
FT   TURN         92     93
FT   STRAND       94     97
FT   STRAND      101    106
SQ   SEQUENCE   107 AA;  11502 MW;  BF938E49 CRC32;

  Query Match             31.5%;  Score 401;  DB 1;  Length 107;
  Best Local Similarity   76.0%;  Pred. No. 1.2e-24;
  Matches   79;  Conservative   12;  Mismatches   11;  Indels    2;  Gaps    2;

Qy      3 LTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRFSG 62
          ||||||| :|| |:|||:|||||||||::||||||| |||| |:|: |||||||| ||||
Db      4 LTQSPAITAASLGQKVTITCSASSSVSSLHWYQQKSGTSPKPWIYEISKLASGVPARFSG 63

Qy     63 SGSGNSYSLTISSMEAEDAATYYCYQGSGYP-FTFGSGTKLEIK 105
          |||| ||||||::||||||| ||| | : ||  |||:|||||:|
Db     64 SGSGTSYSLTINTMEAEDAAIYYCQQWT-YPLITFGAGTKLELK 106


RESULT  14
HV07_MOUSE
ID   HV07_MOUSE     STANDARD;      PRT;   139 AA.
AC   P01751; P01752;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN PRECURSOR V REGION (B1-8 / 186-2).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RX   MEDLINE; 81234548.
RA   BOTHWELL A.L.M., PASKIND M., RETH M., IMANISHI-KARI T., RAJEWSKY K.,
RA   BALTIMORE D.;
RL   CELL 24:625-637(1981).
CC   -!- THE B1-8 MU CHAIN MRNA WAS CLONED FROM A HYBRIDOMA MAKING
CC       ANTIBODIES TO THE HAPTEN (4-HYDROXY-3-NITROPHENYL)ACETYL (NPB
CC       ANTIBODIES).
DR   EMBL; J00529; G195115; -.
DR   PIR; A02034; MHMS18.
DR   HSSP; P01810; 1JHL.
KW   IMMUNOGLOBULIN V REGION; SIGNAL.
FT   SIGNAL        1     19
FT   CHAIN        20    139       IG HEAVY CHAIN V REGION (B1-8 / 186-2).
FT   DOMAIN       20     49       FRAMEWORK 1.
FT   DOMAIN       50     54       COMPLEMENTARITY-DETERMINING 1.
FT   DOMAIN       55     68       FRAMEWORK 2.
FT   DOMAIN       69     85       COMPLEMENTARITY-DETERMINING 2.
FT   DOMAIN       86    117       FRAMEWORK 3.
FT   DOMAIN      118    124       D SEGMENT.
FT   DOMAIN      125    139       JH2 SEGMENT.
FT   DISULFID     41    115       BY SIMILARITY.
FT   NON_TER     139    139
SQ   SEQUENCE   139 AA;  15419 MW;  DEB2C7DA CRC32;
```

```
  Query Match             31.5%;  Score 400;  DB 1;  Length 139;
  Best Local Similarity   58.6%;  Pred. No. 1.9e-24;
  Matches   78;  Conservative   20;  Mismatches   31;  Indels    4;  Gaps    2;

Qy    107 STSGSGKSSEGKGVQLQQSGVELVRGGALVKLSCKASDFNIKDYYIHWVKQRPEQGLEWI 166
          : : :|  |:   ||||| | |||: || |||||||| :    |::||||||| :|||||
Db     11 AATATGVHSQ---VQLQQPGAELVKPGASVKLSCKASGYTFTSYWMHWVKQRPGRGLEWI 67

Qy    167 GWIHPENGNTVYDPKFQGKASITADTSSNAAYLQLSSLTSEDTAVYYCASYYYYSAYYAM 226
          | | | :| | |: ||: ||::| |  |: ||:|||||||||:|||||| | || : |
Db     68 GRIDPNSGGTKYNEKFKSKATLTVDKPSSTAYMQLSSLTSEDSAVYYCARYDYYGSSY-F 126

Qy    227 YYWGQGTSVTVSS 239
           ||||||::||||
Db    127 DYWGQGTTLTVSS 139


RESULT  15
KV4A_MOUSE
ID   KV4A_MOUSE     STANDARD;      PRT;   129 AA.
AC   P01680;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
DE   IG KAPPA CHAIN PRECURSOR V-IV REGION (S107B).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 82115300.
RA   KWAN S.-P., MAX E.E., SEIDMAN J.G., LEDER P., SCHARFF M.D.;
RL   CELL 26:57-66(1981).
CC   -!- THIS PROTEIN, IN WHICH THERE IS A DELETION OF TWO AMINO ACIDS AT
CC       THE V-J RECOMBINATION SITE (AFTER POSITION 118), IS SYNTHESIZED
CC       BUT NOT SECRETED IN CELLS THAT EXPRESS AND SECRETE THE NORMAL
CC       KAPPA CHAIN S107.
DR   EMBL; J00577; G196764; -.
DR   EMBL; V00780; E1181574; -.
DR   PIR; A01943; KVMS7B.
DR   HSSP; P01679; 1FIG.
KW   IMMUNOGLOBULIN V REGION; SIGNAL.
FT   SIGNAL        1     22
FT   CHAIN        23    129       IG KAPPA CHAIN V-IV REGION (S107B).
FT   DOMAIN       23     45       FRAMEWORK 1.
FT   DOMAIN       46     57       COMPLEMENTARITY-DETERMINING 1.
FT   DOMAIN       58     72       FRAMEWORK 2.
FT   DOMAIN       73     79       COMPLEMENTARITY-DETERMINING 2.
FT   DOMAIN       80    111       FRAMEWORK 3.
FT   DOMAIN      112    118       COMPLEMENTARITY-DETERMINING 3.
FT   DOMAIN      119    128       FRAMEWORK 4.
FT   DISULFID     45    111       BY SIMILARITY.
FT   NON_TER     129    129
SQ   SEQUENCE   129 AA;  13833 MW;  8D4D8311 CRC32;


  Query Match             31.1%;  Score 395;  DB 1;  Length 129;
  Best Local Similarity   78.1%;  Pred. No. 4.3e-24;
  Matches   82;  Conservative    7;  Mismatches   12;  Indels    4;  Gaps    2;

Qy      3 LTQSPAIMSASPGEKVTMTCSASSSVSN--MHWYQQKSSTSPKLWVYDTSKLASGVPGRF 60
          ||||||||:|| |:|||||||||||||:  :|||||||  |||  :: || |||||| ||
Db     26 LTQSPAIMAASLGQKVTMTCSASSSVSSSYLHWYQQKSGASPKPLIHRTSNLASGVPARF 85

Qy     61 SGSGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIK 105
          |||||| ||||||||:|||| ||||| | ||||  ||||||||||
Db     86 SGSGSGTSYSLTISSVEAEDDATYYCQQWSGYP--FGSGTKLEIK 128


Search completed: May 13, 1999, 10:05:48
Job time: 615 sec
```

```
                    GenCore version 4.5
           Copyright (c) 1993 - 1998  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         May 13, 1999, 10:05:19 ; Search time 28.48 Seconds
                                         (without alignments)
                                         464.909 Million cell updates/sec

Title:          US-08-704-178-4
Perfect score:  1271
Sequence:       1 MQLTQSPAIMSASPGEKVTM..........SAYYAMYYWGQGTSVTVSSY 240

Scoring table:  BLOSUM62

Searched:       180763 seqs, 55169189 residues

Database :      SPTREMBL_8:*
                1:  sp_fungi:*
                2:  sp_human:*
                3:  sp_invertebrate:*
                4:  sp_mammal:*
                5:  sp_mhc:*
                6:  sp_organelle:*
                7:  sp_phage:*
                8:  sp_plant:*
                9:  sp_bacteria:*
                10:  sp_rodent:*
                11:  sp_virus:*
                12:  sp_vertebrate:*
                13:  sp_unclassified:*
                14:  sp_archea:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                  SUMMARIES
                   %
Result            Query
  No.    Score    Match Length DB  ID                           Description
-------------------------------------------------------------------------------
    1      424     33.4    130 10  P80913                       P80913 mus musculu
    2      212     16.7    100  2  Q15535                       Q15535 homo sapien
    3      188     14.8     85  2  O75724                       O75724 homo sapien
    4      182     14.3     77  2  O75741                       O75741 homo sapien
    5    181.5     14.3     81  2  O75736                       O75736 homo sapien
    6      181     14.2     81  2  O75734                       O75734 homo sapien
    7      176     13.8     77  2  O75726                       O75726 homo sapien
    8    175.5     13.8     75  2  O75743                       O75743 homo sapien
    9      174     13.7     82  2  O75725                       O75725 homo sapien
   10      174     13.7     80  2  O75727                       O75727 homo sapien
   11      174     13.7     77  2  O75728                       O75728 homo sapien
   12      171     13.5     72  2  O75738                       O75738 homo sapien
   13      168     13.2     76  2  O75733                       O75733 homo sapien
   14      167     13.1     86  2  O75740                       O75740 homo sapien
   15      166     13.1     78  2  O75739                       O75739 homo sapien
   16      163     12.8     81  2  O75719                       O75719 homo sapien
   17      163     12.8     81  2  O75721                       O75721 homo sapien
   18    162.5     12.8     82  2  O75732                       O75732 homo sapien
   19      162     12.7     88  2  O75737                       O75737 homo sapien
   20      162     12.7     76  2  O75742                       O75742 homo sapien
   21    160.5     12.6     82  2  O75729                       O75729 homo sapien
   22      158     12.4     78  2  O75730                       O75730 homo sapien
   23      158     12.4     79  2  O75731                       O75731 homo sapien
   24      157     12.4     80  2  O75735                       O75735 homo sapien
   25      155     12.2     74  2  O75744                       O75744 homo sapien
   26      150     11.8     78  2  O75723                       O75723 homo sapien
   27      147     11.6     86  2  O75722                       O75722 homo sapien
   28    140.5     11.1     97  2  O43234                       O43234 homo sapien
   29    135.5     10.7    137  5  Q31181                       Q31181 mus musculu
```

```
   30      135     10.6    509 10  P97710                       P97710 rattus norv
   31      134     10.5    258 12  Q90556                       Q90556 ginglymosto
   32      134     10.5    250 12  Q90569                       Q90569 ginglymosto
   33    133.5     10.5    398  2  O00241                       O00241 homo sapien
   34      133     10.5    158 12  Q90531                       Q90531 ginglymosto
   35      132     10.4     78  2  O75720                       O75720 homo sapien
   36    131.5     10.3     50  2  Q15533                       Q15533 homo sapien
   37    131.5     10.3    503  2  P78324                       P78324 homo sapien
   38    131.5     10.3    258 12  Q90559                       Q90559 ginglymosto
   39    130.5     10.3    145  2  Q16237                       Q16237 homo sapien
   40      130     10.2    418 10  O70426                       O70426 rattus norv
   41    129.5     10.2    237 12  Q90545                       Q90545 ginglymosto
   42    129.5     10.2    254 12  Q90557                       Q90557 ginglymosto
   43      127     10.0    133  5  Q31178                       Q31178 mus musculu
   44    125.5      9.9    509 10  O08907                       O08907 mus musculu
   45    125.5      9.9    509 10  O35924                       O35924 mus musculu

                               ALIGNMENTS


RESULT   1
P80913
ID   P80913      PRELIMINARY;      PRT;   130 AA.
AC   P80913;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   IG KAPPA CHAIN V REGION PRECURSOR.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   JANON K.;
RL   SUBMITTED (SEP-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; X79906; E269393; ALT_TERM.
KW   IMMUNOGLOBULIN V REGION; SIGNAL.
FT   SIGNAL         1     22       POTENTIAL.
FT   CHAIN         23    130       IG KAPPA CHAIN V REGION.
SQ   SEQUENCE   130 AA;  14124 MW;  1DC6F7D9 CRC32;


  Query Match             33.4%;  Score 424;  DB 10;  Length 130;
  Best Local Similarity   79.0%;  Pred. No. 9.6e-26;
  Matches   83;  Conservative    8;  Mismatches   12;  Indels    2;  Gaps    1;

Qy        3 LTQSPAIMSASPGEKVTMTCSASSSVSN--MHWYQQKSSTSPKLWVYDTSKLASGVPGRF 60
            ||||||||||| ||:|||||:||||||:  :||||||  :|||||:| || |||||| ||
Db       26 LTQSPAIMSASLGERVTMTCTASSSVSSSYLHWYQQKPGSSPKLWIYSTSNLASGVPARF 85

Qy       61 SGSGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIK 105
            |||||| |||||||||||||||||||:|    :| ||| |||||||
Db       86 SGSGSGTSYSLTISSMEAEDAATYYCHQYHRFPHTFGGGTKLEIK 130


RESULT   2
Q15535
ID   Q15535      PRELIMINARY;      PRT;   100 AA.
AC   Q15535;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   V KAPPA (FRAGMENT).
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   KATO S., TACHIBANA K., TAKAYAMA N., KATAOKA H., YOSHIDA M.C.,
RA   TAKANO T.;
RL   SUBMITTED (SEP-1990) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; D90161; G1262585; -.
DR   PFAM; PF00047; ig; 1.
FT   NON_TER       1      1
FT   NON_TER     100    100
SQ   SEQUENCE   100 AA;  10871 MW;  06A1440D CRC32;

  Query Match             16.7%;  Score 212;  DB 2;  Length 100;
  Best Local Similarity   42.3%;  Pred. No. 1.1e-09;
  Matches   41;  Conservative   19;  Mismatches   31;  Indels    6;  Gaps    1;

Qy      3 LTQSPAIMSASPGEKVTMTCSASSSV------SNMHWYQQKSSTSPKLWVYDTSKLASGV 56
          :||:|  :| :||:  :::| :  |:      : :: | ||   ||:| :|: |   |||
Db      4 MTQTPLSLSVTPGQPASISCKSIQSLLHSDGKTYLYXYLQKPGQSPQLLIYEVSSRFSGV 63

Qy     57 PGRFSGSGSGNSYSLTISSMEAEDAATYYCYQGSGYP 93
          | ||||||||  ::| || :||||   ||| ||   |
Db     64 PDRFSGSGSGTDFTLKISRVEAEDVGVYYCMQGIHLP 100


RESULT   3
O75724
ID   O75724      PRELIMINARY;      PRT;     85 AA.
AC   O75724;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN VARIABLE REGION (FRAGMENT).
GN   VH.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=INTESTINE;
RA   FISCHER M., KUEPPERS R.;
RT   "Human IgA and IgM secreting intestinal plasma cells carry heavily
RT   mutated VH region genes.";
RL   SUBMITTED (JUL-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AJ009521; E1311442; -.
FT   NON_TER       1      1
FT   NON_TER      85     85
SQ   SEQUENCE   85 AA;  9624 MW;  F1E1DCCC CRC32;

  Query Match             14.8%;  Score 188;  DB 2;  Length 85;
  Best Local Similarity   44.0%;  Pred. No. 6e-08;
  Matches   37;  Conservative   17;  Mismatches   24;  Indels    6;  Gaps    2;

Qy    151 YIHWVKQRPEQGLEWIGWIHPENGNTVYDPKFQGKASITADTSSNAAYLQLSSLTSEDTA 210
          |  |::| | :||||||:|:  :||| |:|  : : :|: ||| |   |:|:|:|: |||
Db      3 YWTWIRQPPGKGLEWIGFIY-YSGNTNYNPSLKSRLTISVDTSKNQFSLKLTSMTAADTA 61

Qy    211 VYYCA-----SYYYYSAYYAMYYW 229
          |||||       |   :|  : ||
Db     62 VYYCARGGGWGSYGSGSYRPLDYW 85


RESULT   4
O75741
ID   O75741      PRELIMINARY;      PRT;     77 AA.
AC   O75741;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN VARIABLE REGION (FRAGMENT).
GN   VH.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=INTESTINE;
RA   FISCHER M., KUEPPERS R.;
RT   "Human IgA and IgM secreting intestinal plasma cells carry heavily
RT   mutated VH region genes.";
RL   SUBMITTED (JUL-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AJ009540; E1311476; -.
FT   NON_TER       1      1
FT   NON_TER      77     77
SQ   SEQUENCE   77 AA;  8734 MW;  1F7F9E8E CRC32;

  Query Match             14.3%;  Score 182;  DB 2;  Length 77;
  Best Local Similarity   46.1%;  Pred. No. 1.5e-07;
  Matches   35;  Conservative   13;  Mismatches   26;  Indels    2;  Gaps    1;

Qy    154 WVKQRPEQGLEWIGWIHPENGNTVYDPKFQGKASITADTSSNAAYLQLSSLTSEDTAVYY 213
          ||:| | :||||:  |      | |    :|: :|: | | |   |||:|:|:|||||||
Db      4 WVRQAPGKGLEWVSGISGSGDYTYYADSVKGRFTISRDNSKNTLSLQLNSVTAEDTAVYY 63

Qy    214 CASYYYYSAYYAMYYW 229
          || :|| |: :   ||
Db     64 CARHYYDSSPFD--YW 77


RESULT   5
O75736
ID   O75736      PRELIMINARY;      PRT;     81 AA.
AC   O75736;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN VARIABLE REGION (FRAGMENT).
GN   VH.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=INTESTINE;
RA   FISCHER M., KUEPPERS R.;
RT   "Human IgA and IgM secreting intestinal plasma cells carry heavily
RT   mutated VH region genes.";
RL   SUBMITTED (JUL-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AJ009534; E1311466; -.
FT   NON_TER       1      1
FT   NON_TER      81     81
SQ   SEQUENCE   81 AA;  9375 MW;  C233614E CRC32;

  Query Match             14.3%;  Score 181.5;  DB 2;  Length 81;
  Best Local Similarity   45.1%;  Pred. No. 1.8e-07;
  Matches   37;  Conservative   13;  Mismatches   25;  Indels    7;  Gaps    3;

Qy    152 IHWVKQRPEQGLEWIGWIHPENGN--TVYDPKFQGKASITADTSSNAAYLQLSSLTSEDT 209
          ||||:| | :||||:  :  :: |  |      :|: :|: | | | |:||: ||  |||
Db      3 IHWVRQAPGKGLEWVARMKRKSDNYATAIAASVRGRFTISRDDSKNMAFLQMDSLKKEDT 62

Qy    210 AVYYCASY--YYYSAYYAMYYW 229
          ||||| :|  ||   || |  |
Db     63 AVYYCVAYEGYY---YYGMDAW 81


RESULT   6
O75734
ID   O75734      PRELIMINARY;      PRT;     81 AA.
AC   O75734;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN VARIABLE REGION (FRAGMENT).
GN   VH.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   TISSUE=INTESTINE;
RA   FISCHER M., KUEPPERS R.;
RT   "Human IgA and IgM secreting intestinal plasma cells carry heavily
RT   mutated VH region genes.";
RL   SUBMITTED (JUL-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AJ009531; E1311462; -.
FT   NON_TER      1      1
FT   NON_TER     81     81
SQ   SEQUENCE   81 AA;  9040 MW;  56B4D902 CRC32;

  Query Match             14.2%;  Score 181;  DB 2;  Length 81;
  Best Local Similarity   42.5%;  Pred. No. 1.9e-07;
  Matches   34;  Conservative   13;  Mismatches   33;  Indels    0;  Gaps    0;

Qy     150 YYIHWVKQRPEQGLEWIGWIHPENGNTVYDPKFQGKASITADTSSNAAYLQLSSLTSEDT 209
           | ::||:| | :||||:  |   : |  |    ||: :|: | : |: |||:| | :|||
Db       2 YSMNWVRQAPGKGLEWVSSISSTSSNIFYGDAVQGRFAISRDNAKNSLYLQISGLGAEDT 61

Qy     210 AVYYCASYYYYSAYYAMYYW 229
           ||||||    |     |  |
Db      62 AVYYCARVLNYKYPNGMDVW 81


RESULT   7
O75726
ID   O75726      PRELIMINARY;      PRT;    77 AA.
AC   O75726;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN VARIABLE REGION (FRAGMENT).
GN   VH.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=INTESTINE;
RA   FISCHER M., KUEPPERS R.;
RT   "Human IgA and IgM secreting intestinal plasma cells carry heavily
RT   mutated VH region genes.";
RL   SUBMITTED (JUL-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AJ009523; E1311446; -.
FT   NON_TER      1      1
FT   NON_TER     77     77
SQ   SEQUENCE   77 AA;  9022 MW;  EB7B458D CRC32;

  Query Match             13.8%;  Score 176;  DB 2;  Length 77;
  Best Local Similarity   42.9%;  Pred. No. 4.4e-07;
  Matches   30;  Conservative   17;  Mismatches   23;  Indels    0;  Gaps    0;

Qy     150 YYIHWVKQRPEQGLEWIGWIHPENGNTVYDPKFQGKASITADTSSNAAYLQLSSLTSEDT 209
           |:: ||:| | :||||:  |: : |   |    :|: :|: | ::|: |||::|| :|||
Db       1 YFLTWVRQAPGKGLEWVATINEDGGEKYYVDSVKGRFTISRDDATNSLYLQMNSLRAEDT 60

Qy     210 AVYYCASYYY 219
           |||||    |
Db      61 AVYYCVRDSY 70


RESULT   8
O75743
ID   O75743      PRELIMINARY;      PRT;    75 AA.
AC   O75743;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN VARIABLE REGION (FRAGMENT).
GN   VH.
OS   HOMO SAPIENS (HUMAN).
```

```
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=INTESTINE;
RA   FISCHER M., KUEPPERS R.;
RT   "Human IgA and IgM secreting intestinal plasma cells carry heavily
RT   mutated VH region genes.";
RL   SUBMITTED (JUL-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AJ009543; E1311482; -.
FT   NON_TER      1      1
FT   NON_TER     75     75
SQ   SEQUENCE   75 AA;  8667 MW;  8D5C330F CRC32;

  Query Match             13.8%;  Score 175.5;  DB 2;  Length 75;
  Best Local Similarity   47.9%;  Pred. No. 4.7e-07;
  Matches   34;  Conservative   14;  Mismatches   22;  Indels    1;  Gaps    1;

Qy     151 YIHWVKQRPEQGLEWIGWIHPENGNTVYDPKFQGKASITADTSSNAAYLQLSSLTSEDTA 210
           |  |::| | :||||||  |   | | |:|  : : :|:|||| |   |:::|:|: |||
Db       1 YWAWIRQPPGKGLEWIGG-HYYPGGTYYNPSLKSRVTISADTSKNHFSLKMTSVTAADTA 59

Qy     211 VYYCASYYYYS 221
           |||||  : ||
Db      60 VYYCARRHGYS 70


RESULT   9
O75725
ID   O75725      PRELIMINARY;      PRT;    82 AA.
AC   O75725;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN VARIABLE REGION (FRAGMENT).
GN   VH.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=INTESTINE;
RA   FISCHER M., KUEPPERS R.;
RT   "Human IgA and IgM secreting intestinal plasma cells carry heavily
RT   mutated VH region genes.";
RL   SUBMITTED (JUL-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AJ009522; E1311444; -.
FT   NON_TER      1      1
FT   NON_TER     82     82
SQ   SEQUENCE   82 AA;  9566 MW;  1F2E1379 CRC32;

  Query Match             13.7%;  Score 174;  DB 2;  Length 82;
  Best Local Similarity   40.7%;  Pred. No. 6.8e-07;
  Matches   33;  Conservative   18;  Mismatches   28;  Indels    2;  Gaps    2;

Qy     150 YYIHWVKQRPEQGLEWIGWIHPENGNTVYDPKFQGKASITADTSSNAAYLQLSSLTSEDT 209
           :|  |::| | :|||||| ::  :|   |:|  | : :|: ||| |   |:|:|:|: ||
Db       3 FYWGWIRQPPGKGLEWIGTVY-YSGTNYYNPSLQSRVTISVDTSKNQFSLRLTSVTAADT 61

Qy     210 AVYYCASYYYYSAYYAMY-YW 229
           ||||||   :    :  : ||
Db      62 AVYYCARPFEGGRAWPYFDYW 82


RESULT  10
O75727
ID   O75727      PRELIMINARY;      PRT;    80 AA.
AC   O75727;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
```

```
DE   IG HEAVY CHAIN VARIABLE REGION (FRAGMENT).
GN   VH.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=INTESTINE;
RA   FISCHER M., KUEPPERS R.;
RT   "Human IgA and IgM secreting intestinal plasma cells carry heavily
RT   mutated VH region genes.";
RL   SUBMITTED (JUL-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AJ009524; E1311448; -.
FT   NON_TER      1      1
FT   NON_TER     80     80
SQ   SEQUENCE   80 AA;  9351 MW;  63DE158A CRC32;


  Query Match             13.7%;  Score 174;  DB 2;  Length 80;
  Best Local Similarity   44.7%;  Pred. No. 6.6e-07;
  Matches   34;  Conservative   16;  Mismatches   24;  Indels    2;  Gaps    1;

Qy    151 YIHWVKQRPEQGLEWIGWIHPENGN--TVYDPKFQGKASITADTSSNAAYLQLSSLTSED 208
          |: ||:| | :||||:| |  :  :  | |    :|: :|: | | |: |||::|| ::|
Db      1 YMDWVRQAPGKGLEWVGRIRKKANSYTTEYAASVKGRFTISRDDSKNSLYLQMNSLKTDD 60

Qy    209 TAVYYCASYYYYSAYY 224
          |||||||  :  || |
Db     61 TAVYYCARDHLRSAAY 76


RESULT  11
075728
ID   075728      PRELIMINARY;      PRT;    77 AA.
AC   075728;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN VARIABLE REGION (FRAGMENT).
GN   VH.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=INTESTINE;
RA   FISCHER M., KUEPPERS R.;
RT   "Human IgA and IgM secreting intestinal plasma cells carry heavily
RT   mutated VH region genes.";
RL   SUBMITTED (JUL-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AJ009525; E1311450; -.
FT   NON_TER      1      1
FT   NON_TER     77     77
SQ   SEQUENCE   77 AA;  8925 MW;  EA14CFDF CRC32;


  Query Match             13.7%;  Score 174;  DB 2;  Length 77;
  Best Local Similarity   43.9%;  Pred. No. 6.3e-07;
  Matches   29;  Conservative   15;  Mismatches   22;  Indels    0;  Gaps    0;

Qy    150 YYIHWVKQRPEQGLEWIGWIHPENGNTVYDPKFQGKASITADTSSNAAYLQLSSLTSEDT 209
          | ::|::| | :||:|:  :    | | |    :|: ||: | | |  |||::|| :|||
Db      1 YVMNWIRQAPGEGLQWVSTVTENGGVTFYPDSVKGRFSISRDNSRNTLYLQMNSLRAEDT 60

Qy    210 AVYYCA 215
          |:||||
Db     61 ALYYCA 66


RESULT  12
075738
ID   075738      PRELIMINARY;      PRT;    72 AA.
AC   075738;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN VARIABLE REGION (FRAGMENT).
GN   VH.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=INTESTINE;
RA   FISCHER M., KUEPPERS R.;
RT   "Human IgA and IgM secreting intestinal plasma cells carry heavily
RT   mutated VH region genes.";
RL   SUBMITTED (JUL-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AJ009537; E1311470; -.
FT   NON_TER      1      1
FT   NON_TER     72     72
SQ   SEQUENCE   72 AA;  8345 MW;  E0FB044A CRC32;


  Query Match             13.5%;  Score 171;  DB 2;  Length 72;
  Best Local Similarity   39.7%;  Pred. No. 9.9e-07;
  Matches   31;  Conservative   16;  Mismatches   23;  Indels    8;  Gaps    1;

Qy    152 IHWVKQRPEQGLEWIGWIHPENGNTVYDPKFQGKASITADTSSNAAYLQLSSLTSEDTAV 211
          ::||:| | :||||| :|    |   |    :|: :|::| : |: |||::||  :||||
Db      3 MNWVRQAPGKGLEWISYIRSSTGFIYYTDSVRGRFTISSDIAKNSVYLQMNSLRDDDTAV 62

Qy    212 YYCASYYYYSAYYAMYYW 229
          ||||          : ||
Db     63 YYCAR--------DLAYW 72


RESULT  13
075733
ID   075733      PRELIMINARY;      PRT;    76 AA.
AC   075733;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN VARIABLE REGION (FRAGMENT).
GN   VH.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=INTESTINE;
RA   FISCHER M., KUEPPERS R.;
RT   "Human IgA and IgM secreting intestinal plasma cells carry heavily
RT   mutated VH region genes.";
RL   SUBMITTED (JUL-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AJ009530; E1311460; -.
FT   NON_TER      1      1
FT   NON_TER     76     76
SQ   SEQUENCE   76 AA;  8543 MW;  C95EB66A CRC32;


  Query Match             13.2%;  Score 168;  DB 2;  Length 76;
  Best Local Similarity   43.8%;  Pred. No. 1.8e-06;
  Matches   35;  Conservative   12;  Mismatches   25;  Indels    8;  Gaps    2;

Qy    152 IHWVKQRPEQGLEWIGWIHPENGN--TVYDPKFQGKASITADTSSNAAYLQLSSLTSEDT 209
          ||||:|   :||||:  |  :  :  | |    :|:  :: | | | |||::|| ||||
Db      3 IHWVRQTSGKGLEWVARIRSKTHSYATAYAASVEGRFVVSRDDSKNTAYLQMNSLKSEDT 62

Qy    210 AVYYCASYYYYSAYYAMYYW 229
          ||||||      |  :  ||
Db     63 AVYYCA------AGSSSEYW 76


RESULT  14
```

```
O75740
ID   O75740      PRELIMINARY;      PRT;    86 AA.
AC   O75740;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN VARIABLE REGION (FRAGMENT).
GN   VH.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=INTESTINE;
RA   FISCHER M., KUEPPERS R.;
RT   "Human IgA and IgM secreting intestinal plasma cells carry heavily
RT   mutated VH region genes.";
RL   SUBMITTED (JUL-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AJ009539; E1311474; -.
FT   NON_TER       1      1
FT   NON_TER      86     86
SQ   SEQUENCE   86 AA;  9625 MW;  07627E8C CRC32;


  Query Match             13.1%;  Score 167;  DB 2;  Length 86;
  Best Local Similarity   45.6%;  Pred. No. 2.4e-06;
  Matches   31;  Conservative   15;  Mismatches   20;  Indels    2;  Gaps    1;

Qy    151 YIHWVKQRPEQGLEWIGWI--HPENGNTVYDPKFQGKASITADTSSNAAYLQLSSLTSED 208
          :::||:| | :||||:| |    : | | |    :|: :|: | | |  |||::|| :||
Db      3 WMNWVRQAPGKGLEWVGRIKSKTDGGTTDYAAPVKGRFTISRDDSKNTLYLQMNSLKTED 62

Qy    209 TAVYYCAS 216
          |||||| :
Db     63 TAVYYCTT 70


RESULT  15
O75739
ID   O75739      PRELIMINARY;      PRT;    78 AA.
AC   O75739;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   IG HEAVY CHAIN VARIABLE REGION (FRAGMENT).
GN   VH.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=INTESTINE;
RA   FISCHER M., KUEPPERS R.;
RT   "Human IgA and IgM secreting intestinal plasma cells carry heavily
RT   mutated VH region genes.";
RL   SUBMITTED (JUL-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AJ009538; E1311472; -.
FT   NON_TER       1      1
FT   NON_TER      78     78
SQ   SEQUENCE   78 AA;  9031 MW;  05D273B9 CRC32;


  Query Match             13.1%;  Score 166;  DB 2;  Length 78;
  Best Local Similarity   40.5%;  Pred. No. 2.6e-06;
  Matches   32;  Conservative   20;  Mismatches   23;  Indels    4;  Gaps    2;

Qy    151 YIHWVKQRPEQGLEWIGWIHPENGNTVYDPKFQGKASITADTSSNAAYLQLSSLTSEDTA 210
          |  |::| | :||||||:|:   |:  |:|  : : :|:  || |   | |:|:|: |||
Db      1 YWSWIRQSPGKGLEWIGYIY-YTGSINYNPSLKSRVTISLGTSENQFSLNLTSVTAADTA 59

Qy    211 VYYCASYYYYSAYYAMYYW 229
          ||:||    :|: : :||:
Db     60 VYFCAR---HSSSWPIYYF 75
```

```
Search completed: May 13, 1999, 10:05:19
Job time: 657 sec
```

```
                     GenCore version 4.5
           Copyright (c) 1993 - 1998  Compugen Ltd.


OM protein - protein search, using sw model

Run on:         May 13, 1999, 10:02:41 ; Search time 23.53 Seconds
                                         (without alignments)
                                         206.293 Million cell updates/sec

Title:          US-08-704-178-4
Perfect score:  1271
Sequence:       1 MQLTQSPAIMSASPGEKVTM..........SAYYAMYYWGQGTSVTVSSY 240

Scoring table:  BLOSUM62

Searched:       162890 seqs, 20225328 residues

Database :      A_Geneseq_34:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                    SUMMARIES
                  %
Result          Query
  No.   Score   Match Length DB  ID                 Description
-----------------------------------------------------------------------
    1    1264   99.4    239  1  W15186              Single-chain anti-
    2    1259   99.1    238  1  R45443              Sequence of the si
    3   829.5   65.3    244  1  R60780              Fv(GP-2) immunosup
    4   821.5   64.6    241  1  R12798              OVB3 light and hea
    5   795.5   62.6    237  1  R94020              Anti-erbB2 scFv. I
    6   795.5   62.6    237  1  W15185              Single-chain anti-
    7   790.5   62.2    236  1  R45442              Sequence of the si
    8     763   60.0    240  1  R15443              Single chain Fv fr
    9     759   59.7    240  1  W25784              Anti-gp130 antibod
   10     752   59.2    553  1  W11508              Single chain anti-
   11     752   59.2    553  1  W73223              H22-anti-CEA antib
   12   751.5   59.1    239  1  W02191              18-2-3/TRY202' sin
   13   748.5   58.9    239  1  R43679              Single chain polyp
   14   748.5   58.9    239  1  R99649              Single chain bindi
   15   744.5   58.6    241  1  R06482              18-2-3-/TRY202'. S
   16   727.5   57.2    245  1  W25783              Anti-T3 antibody d
   17   725.5   57.1    473  1  W26646              Chimeric receptor
   18   725.5   57.1    514  1  W26647              Chimeric receptor
   19   725.5   57.1    403  1  W26648              Chimeric receptor
   20   725.5   57.1    651  1  W26649              Chimeric receptor
   21   725.5   57.1    692  1  W26650              Chimeric receptor
   22   725.5   57.1    382  1  W26651              Chimeric receptor
   23     724   57.0    250  1  R43678              Single chain polyp
   24     724   57.0    250  1  R99648              Single chain bindi
   25     724   57.0    250  1  W02190              TRY104b single cha
   26     719   56.6    246  1  R43677              Single chain polyp
   27     715   56.3    250  1  R06481              TRY104b. Single po
   28     715   56.3    246  1  R99647              Single chain bindi
   29     714   56.2    486  1  R37649              Sequence of divale
   30   712.5   56.1    271  1  R20185              Sequence of the sh
   31     712   56.0    262  1  R97380              CC49/212 SCA PLAP
   32     712   56.0    264  1  R97381              PLAP CC49/212 SCA
   33     711   55.9    242  1  R43680              Single chain polyp
   34     711   55.9    242  1  R99650              Single chain bindi
   35     711   55.9    242  1  W02192              18-2-3/TRY59 singl
   36     701   55.2    242  1  R06483              18-2-3-/TRY59. Sin
   37   700.5   55.1    360  1  W35128              R. pipiens recombi
   38   700.5   55.1    358  1  W35127              R. pipiens recombi
   39   700.5   55.1    358  1  W35130              R. pipiens recombi
   40     699   55.0    246  1  W02189              TRY59 single chain
   41   698.5   55.0    379  1  W35126              R. pipiens recombi
   42   698.5   55.0    355  1  W35129              R. pipiens recombi
   43   698.5   55.0    355  1  W35125              R. pipiens recombi
```

```
   44     696   54.8    301  1  W37085              Anti-human SC sing
   45     691   54.4    243  1  R60781              Fv(GP-4) immunosup


                                  ALIGNMENTS


RESULT   1
W15186
ID   W15186 standard; Protein; 239 AA.
AC   W15186;
DT   05-JUN-1997  (first entry)
DE   Single-chain anti-erbB2 antibody e21(Fv).
KW   Single chain antibody; variable region; light chain; heavy chain;
KW   breast cancer; ovarian cancer; non-small cell lung carcinoma;
KW   immunodiagnosis; treatment; cytotoxic agent; erbB-2.
OS   Mus musculus.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   region          106. .119
FT                   /label= linker
PN   US5587458-A.
PD   24-DEC-1996.
PF   07-OCT-1991; 772270.
PR   07-OCT-1991; US-772270.
PR   30-JUN-1992; US-906555.
PR   14-MAY-1993; US-061092.
PA   (ARON-) ARONEX PHARM INC.
PI   Bird RE,  Kasprzyk PG,  King CR;
DR   WPI; 97-064831/06.
DR   N-PSDB; T65007.
PT   Single chain antibodies specific for erbB-2 protein, gp185  - with
PT   labels or cytotoxin, useful for detection and treatment of tumour
PT   cells expressing this protein
PS   Claim 2; Columns 27-30; 28pp; English.
CC   The present sequence represents a claimed single-chain antibody,
CC   designated e21(Fv), which binds to erbB-2. Monoclonal antibody e21
CC   was generated by immunising mice with N/erbB-2 cells overexpressing
CC   the gp185 protein, removing spleen cells and producing hybridomas
CC   by standard techniques. Messenger RNA coding for the anti-erbB-2
CC   monoclonal antibody was isolated and converted to cDNA. Regions
CC   coding for the heavy- and light- chain variable regions were then
CC   amplified by PCR and joined via a sequence encoding a peptide
CC   linker. The resulting single-chain antibody is useful for in vitro
CC   diagnosis of tumour cells which overexpress the erbB-2 gp185
CC   marker, e.g. breast, ovarian and non-small cell lung carcinomas,
CC   and, when coupled to a cytotoxic agent, to treat such tumours.
SQ   Sequence   239 AA;


  Query Match             99.4%;  Score 1264;  DB 1;  Length 239;
  Best Local Similarity   100.0%;  Pred. No. 4.5e-82;
  Matches  239;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 MQLTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRF 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 MQLTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRF 60

Qy       61 SGSGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIKGSTSGSGKSSEGKGV 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 SGSGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIKGSTSGSGKSSEGKGV 120

Qy      121 QLQQSGVELVRGGALVKLSCKASDFNIKDYYIHWVKQRPEQGLEWIGWIHPENGNTVYDP 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 QLQQSGVELVRGGALVKLSCKASDFNIKDYYIHWVKQRPEQGLEWIGWIHPENGNTVYDP 180

Qy      181 KFQGKASITADTSSNAAYLQLSSLTSEDTAVYYCASYYYYSAYYAMYYWGQGTSVTVSS 239
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 KFQGKASITADTSSNAAYLQLSSLTSEDTAVYYCASYYYYSAYYAMYYWGQGTSVTVSS 239


RESULT   2
R45443
```

```
ID   R45443 standard; Protein; 238 AA.
AC   R45443;
DT   21-JUL-1994  (first entry)
DE   Sequence of the single chain anti-erbB2 antibody, Ab no.21.
KW   Single chain anti-erbB1 antibody; cancer therapy; prevention;
KW   monoclonal antibody.
OS   Synthetic.
PN   WO9400136-A.
PD   06-JAN-1994.
PF   21-OCT-1992; U08545.
PR   30-JUN-1992; US-906555.
PA   (MOLE-) MOLECULAR ONCOLOGY INC.
PI   Kasprzyk PG,  King CR;
DR   WPI; 94-025878/03.
DR   N-PSDB; Q55181.
PT   Treatment of malignancies over-expressing ERB-[2 - using at least
PT   2 monoclonal antibodies which recognise different epitopes on
PT   gp185
PS   Example; Fig 8; 37pp; English.
CC   The source of human erbB-2 protein for the prodn. of antibodies no.
CC   23 and 21 (Q55180/R45442; Q55181/R45443) is a NIH/3T3 cell
CC   engineered to express the human erbB-2 protein on its surface (N/
CC   erbB-2). Abs no. 21 and 23 are directed against the extracellular
CC   domain of gp185 erbB-2. Nude mice manipulated to produce rapidly
CC   growing tumours were used in a trial of the efficacy of the Abs. In
CC   animals given a combination of the 2 Abs, tumours completely
CC   regressed after 11 days.
SQ   Sequence   238 AA;

  Query Match             99.1%;  Score 1259;  DB 1;  Length 238;
  Best Local Similarity   100.0%;  Pred. No. 1e-81;
  Matches  238;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 QLTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRFS 61
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 QLTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRFS 60

Qy     62 GSGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIKGSTSGSGKSSEGKGVQ 121
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GSGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIKGSTSGSGKSSEGKGVQ 120

Qy    122 LQQSGVELVRGGALVKLSCKASDFNIKDYYIHWVKQRPEQGLEWIGWIHPENGNTVYDPK 181
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 LQQSGVELVRGGALVKLSCKASDFNIKDYYIHWVKQRPEQGLEWIGWIHPENGNTVYDPK 180

Qy    182 FQGKASITADTSSNAAYLQLSSLTSEDTAVYYCASYYYYSAYYAMYYWGQGTSVTVSS 239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 FQGKASITADTSSNAAYLQLSSLTSEDTAVYYCASYYYYSAYYAMYYWGQGTSVTVSS 238

RESULT    3
R60780
ID   R60780 standard; Protein; 244 AA.
AC   R60780;
DT   09-MAY-1995  (first entry)
DE   Fv(GP-2) immunosuppressive.
KW   MAb; monoclonal antibody; hybridoma; interleukin-2; IL-2;
KW   Fv; antibody variable region; GP-2; Fv(GP-2);
KW   immunosuppressive.
OS   Mus sp.
PN   EP-621338-A.
PD   26-OCT-1994.
PF   21-APR-1994; 106257.
PR   21-APR-1993; JP-094491.
PR   07-MAR-1994; JP-036065.
PA   (AJIN ) AJINOMOTO KK.
PI   Hamura J,  Kanayama Y,  Nakazawa H,  Shimamura T,  Sugamura K;
PI   Takeshita T;
DR   WPI; 94-325948/41.
DR   N-PSDB; Q73678.
PT   Immunosuppressant polypeptide - has ability to block
PT   interleukin-2 response
```

```
PS   Claim 8; Page 26-27; 37pp; English.
CC   MAb capable of binding to the gamma chain of the IL-2 receptor, and
CC   thus of blocking the IL-2 response, is produced by mouse hybridoma
CC   line GP-2 (FERM BP-4641).  DNA encoding the variable region of
CC   this MAb was expressed in E. coli, yielding Fv(GP-2) with
CC   immumosuppressive activity.
SQ   Sequence   244 AA;

  Query Match             65.3%;  Score 829.5;  DB 1;  Length 244;
  Best Local Similarity   68.3%;  Pred. No. 1.4e-51;
  Matches  164;  Conservative   27;  Mismatches   46;  Indels    3;  Gaps    2;

Qy      3 LTQSPAIMSASPGEKVTMTCSASSSVSN--MHWYQQKSSTSPKLWVYDTSKLASGVPGRF 60
          |||||:||||| ||:|||||:||||||:  :||||||  :|||||:| || |||||| ||
Db      5 LTQSPSIMSASLGERVTMTCTASSSVSSSYLHWYQQKPGSSPKLWIYSTSNLASGVPARF 64

Qy     61 SGSGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIK-GSTSGSGKSSEGKG 119
          |||||| |||||||||||||||||||:|    | |||:|||||:|   :||||  |:
Db     65 SGSGSGTSYSLTISSMEAEDAATYYCHQYHRSPLTFGAGTKLELKVDKSSGSGSESKSTQ 124

Qy    120 VQLQQSGVELVRGGALVKLSCKASDFNIKDYYIHWVKQRPEQGLEWIGWIHPENGNTVYD 179
          |:|::|| |||| || |||||||| :    |::||||||  ||||||| |:| :|:| ||
Db    125 VKLEESGSELVRPGASVKLSCKASGYTFTSYWMHWVKQRHGQGLEWIGNIYPGSGSTNYD 184

Qy    180 PKFQGKASITADTSSNAAYLQLSSLTSEDTAVYYCASYYYYSAYYAMYYWGQGTSVTVSS 239
           ||: ||::| ||||: ||: ||||||||:|||||        |||| ||||||||||||
Db    185 EKFKSKATLTVDTSSSTAYMHLSSLTSEDSAVYYCTRSSRNWVYYAMDYWGQGTSVTVSS 244

RESULT    4
R12798
ID   R12798 standard; Protein; 241 AA.
AC   R12798;
DT   16-SEP-1991  (first entry)
DE   OVB3 light and heavy chains.
KW   Immunoglobulin; immunotoxin; antibody; variable region; PE40;
KW   pseudomonas exotoxin.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   region          1. .107
FT                   /label= Light chain
FT   region          108. .121
FT                   /label= linker
FT   region          122
FT                   /label= Heavy cahin
FT   region          24. .33
FT                   /label= CDR1
FT   region          49. .54
FT                   /label= CDR2
FT   region          88. .96
FT                   /label= CDR3
FT   region          151. .155
FT                   /label= CDR1
FT   region          170. .186
FT                   /label= CDR2
FT   region          219. .229
FT                   /label= CDR3
FT   peptide         1. .20
FT                   /label= N-terminal sequencing
FT   peptide         121. .134
FT                   /label= N-terminal sequencing
PN   US7463111-A.
PD   28-MAY-1991.
PF   11-JAN-1990; 133066.
PR   11-JAN-1990; US-463111.
PA   (USSH ) NAT INST OF HEALTH.
PI   Pastan I;
DR   WPI; 91-200877/27.
DR   N-PSDB; Q12405.
PT   Rapid cloning of antibody genes as single chain immuno-toxins -
PT   by fusing genes with DNA encoding cytotoxic molecules, to kill
PT   target cells
```

```
PS   Disclosure; Fig 3; 26pp; English.
CC   The sequence comprises the L and H chains of murine MAb OVB3,
CC   joined together via a linker.  The sequence is encoded by clone
CC   pOVB3158-1 in which the 3' end of the DNA is linked to the 5' end
CC   of the coding sequence for PE40, a Pseudomonas exotoxin lacking
CC   domain I.  The protein expressed by the clone can be used to kill
CC   specific target cells.
SQ   Sequence   241 AA;


  Query Match             64.6%;  Score 821.5;  DB 1;  Length 241;
  Best Local Similarity   66.1%;  Pred. No. 5.1e-51;
  Matches  158;  Conservative   31;  Mismatches   47;  Indels    3;  Gaps    3;

Qy      3 LTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRFSG 62
          |||||||||||||||||||||| |:|| |||||||||||||||:|:||||||||||||||
Db      4 LTQSPAIMSASPGEKVTMTCSAISTVSYMHWYQQKSSTSPKLWIYETSKLASGVPGRFSG 63

Qy     63 SGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIK-GSTSGSGKSSEGKGVQ 121
          |||||||||||||||||| |||||:||||||:||| |||||:| | :||||  |:   ||
Db     64 SGSGNSYSLTISSMEAEDVATYYCFQGSGYPYTFGGGTKLELKEGKSSGSGSESKVDDVQ 123

Qy    122 LQQSGVELVRGGALVKLSCKASDFNIKDYYIHWVKQRPEQGLEWIGWIHPENGNTVYDPK 181
          | :||  ||: |   :||| || |    : :|||:| ||:||||: :|    |||:|
Db    124 LVESGGGLVQPGGSRQLSCAASGFTFSSFGMHWVRQAPEKGLEWVAYI-SRGGNTIYYAN 182

Qy    182 -FQGKASITADTSSNAAYLQLSSLTSEDTAVYYCASYYYYSAYYAMYYWGQGTSVTVSS 239
            :|: :|: |    |  :||::|| |:|||:||||  :||  :||| ||||||::| ||
Db    183 TVKGRFTISRDNPKNTLFLQMTSLRSDDTAMYYCARSHYYGYFYAMDYWGQGTTLTGSS 241


RESULT    5
R94020
ID   R94020 standard; Protein; 237 AA.
AC   R94020;
DT   21-MAY-1996  (first entry)
DE   Anti-erbB2 scFv.
KW   Oncoprotein; erbB2; cell proliferation; tumour; cancer;
KW   intracellular antibody homologue; single chain antibody; scFv;
KW   gene therapy.
OS   Synthetic.
PN   WO9607321-A1.
PD   14-MAR-1996.
PF   23-AUG-1995; U10740.
PR   06-SEP-1994; US-301339.
PR   06-JUN-1995; US-468252.
PA   (UABR-) UAB RES FOUND.
PI   Curiel DT,  Deshane J;
DR   WPI; 96-171307/17.
DR   N-PSDB; T17728.
PT   Inhibition of proliferation or survival of, esp. malignant erbB2,
PT   cells - by introducing nucleic acid mol. encoding antibody homologue
PT   which is expressed and binds, pref. erbB2, protein intracellularly
PS   Disclosure; Page 29-30; 48pp; English.
CC   A nucleic acid comprises a first sequence encoding a signal peptide
CC   (R94019) linked to a second sequence (T17728) encoding a single
CC   chain Fv fragment (R94020) that binds a human erbB2 oncoprotein.
CC   The anti-erbB2 sFv portion is obtained by PCR using e23scFv
CC   plasmid as template.  The signal peptide directs the scFv to the
CC   endoplasmic reticulum.  The nucleic acid is incorporated into a
CC   plasmid or viral vector to facilitate expression of the scFv antibody
CC   homologue within e.g. an epithelial carcinoma cell.  Intracellular
CC   expression of the homologue inhibits surface expression of erbB2 and
CC   thereby inhibits cell proliferation and cell survival and decreases
CC   tumorigenicity.
SQ   Sequence   237 AA;


  Query Match             62.6%;  Score 795.5;  DB 1;  Length 237;
  Best Local Similarity   64.3%;  Pred. No. 3.3e-49;
  Matches  155;  Conservative   32;  Mismatches   45;  Indels    9;  Gaps    2;

Qy      1 MQLTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRF 60
          :||||||||:||||||||||||| |: ||| |||||||  :||| |:| || |||||| ||
Db      3 LQLTQSPAILSASPGEKVTMTCRATPSVSYMHWYQQKPGSSPKPWIYTTSNLASGVPARF 62

Qy     61 SGSGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIKGSTSGSGKSSEGKGV 120
          || ||| |||||:| :|||||||||| | |  | ||| |:||||||||||||||||||||
Db     63 SGGGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIKGSTSGSGKSSEGKGV 122

Qy    121 QLQQSGVELVRGGALVKLSCKASDFNIKDYYIHWVKQRPEQGLEWIGWIHPENGNTVYDP 180
          |||:|| |:|: |  :|:||| | ::   : ::||||   : ||||| |:| ||:| |:
Db    123 QLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGDTNYNQ 182

Qy    181 KFQGKASITADTSSNAAYLQLSSLTSEDTAVYYCA---SYYYYSAYYAMYYWGQGTSVTV 237
          ||:|||: | | ||: ||::| ||||||:||||||   : :|:        || ||:|||
Db    183 KFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARRVTDWYFDV------WGAGTTVTV 236

Qy    238 S 238
          |
Db    237 S 237


RESULT    6
W15185
ID   W15185 standard; Protein; 237 AA.
AC   W15185;
DT   05-JUN-1997  (first entry)
DE   Single-chain anti-erbB2 antibody e23(Fv).
KW   Single chain antibody; variable region; light chain; heavy chain;
KW   breast cancer; ovarian cancer; non-small cell lung carcinoma;
KW   immunodiagnosis; treatment; cytotoxic agent; erbB-2.
OS   Mus musculus.
OS   Synthetic.
FH   Key              Location/Qualifiers
FT   region           108. .121
FT                    /label= linker
PN   US5587458-A.
PD   24-DEC-1996.
PF   07-OCT-1991; 772270.
PR   07-OCT-1991; US-772270.
PR   30-JUN-1992; US-906555.
PR   14-MAY-1993; US-061092.
PA   (ARON-) ARONEX PHARM INC.
PI   Bird RE,  Kasprzyk PG,  King CR;
DR   WPI; 97-064831/06.
DR   N-PSDB; T65006.
PT   Single chain antibodies specific for erbB-2 protein, gp185  - with
PT   labels or cytotoxin, useful for detection and treatment of tumour
PT   cells expressing this protein
PS   Claim 2; Columns 25-28; 28pp; English.
CC   The present sequence represents a claimed single-chain antibody,
CC   designated e23(Fv), which binds to erbB-2. Monoclonal antibody e23
CC   was generated by immunising mice with N/erbB-2 cells overexpressing
CC   the gp185 protein, removing spleen cells and producing hybridomas
CC   by standard techniques. Messenger RNA coding for the anti-erbB-2
CC   monoclonal antibody was isolated and converted to cDNA. Regions
CC   coding for the heavy- and light- chain variable regions were then
CC   amplified by PCR and joined via a sequence encoding a peptide
CC   linker. The resulting single-chain antibody is useful for in vitro
CC   diagnosis of tumour cells which overexpress the erbB-2 gp185
CC   marker, e.g. breast, ovarian and non-small cell lung carcinomas,
CC   and, when coupled to a cytotoxic agent, to treat such tumours.
SQ   Sequence   237 AA;


  Query Match             62.6%;  Score 795.5;  DB 1;  Length 237;
  Best Local Similarity   64.3%;  Pred. No. 3.3e-49;
  Matches  155;  Conservative   32;  Mismatches   45;  Indels    9;  Gaps    2;

Qy      1 MQLTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRF 60
          :||||||||:||||||||||||| |: ||| |||||||  :||| |:| || |||||| ||
Db      3 LQLTQSPAILSASPGEKVTMTCRATPSVSYMHWYQQKPGSSPKPWIYTTSNLASGVPARF 62

Qy     61 SGSGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIKGSTSGSGKSSEGKGV 120
          || ||| |||||:| :|||||||||| | |  | ||| |:||||||||||||||||||||
Db     63 SGGGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIKGSTSGSGKSSEGKGV 122

Qy    121 QLQQSGVELVRGGALVKLSCKASDFNIKDYYIHWVKQRPEQGLEWIGWIHPENGNTVYDP 180
          |||:|| |:|: |  :|:||| | ::   : ::||||   : ||||| |:| ||:| |:
Db    123 QLQESGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGDTNYNQ 182

Qy    181 KFQGKASITADTSSNAAYLQLSSLTSEDTAVYYCA---SYYYYSAYYAMYYWGQGTSVTV 237
          ||:|||: | | ||: ||::| ||||||:||||||   : :|:        || ||:|||
Db    183 KFKGKATFTVDKSSSTAYMELLSLTSEDSAVYYCARRVTDWYFDV------WGAGTTVTV 236

Qy    238 S 238
          |
Db    237 S 237


RESULT    7
R45442
ID   R45442 standard; Protein; 236 AA.
AC   R45442;
DT   21-JUL-1994  (first entry)
DE   Sequence of the single chain anti-erbB2 antibody, Ab no.23.
KW   Single chain anti-erbB1 antibody; cancer therapy; prevention;
KW   monoclonal antibody.
OS   Synthetic.
PN   WO9400136-A.
PD   06-JAN-1994.
PF   21-OCT-1992; U08545.
PR   30-JUN-1992; US-906555.
PA   (MOLE-) MOLECULAR ONCOLOGY INC.
PI   Kasprzyk PG,  King CR;
DR   WPI; 94-025878/03.
DR   N-PSDB; Q55180.
PT   Treatment of malignancies over-expressing ERB-[2 - using at least
PT   2 monoclonal antibodies which recognise different epitopes on
PT   gp185
PS   Example; Fig 7; 37pp; English.
CC   The source of human erbB-2 protein for the prodn. of antibodies no.
CC   23 and 21 (Q55180/R45442; Q55181/R45443) is a NIH/3T3 cell
CC   engineered to express the human erbB-2 protein on its surface (N/
CC   erbB-2). Abs no. 21 and 23 are directed against the extracellular
CC   domain of gp185 erbB-2. Nude mice manipulated to produce rapidly
CC   growing tumours were used in a trial of the efficacy of the Abs. In
CC   animals given a combination of the 2 Abs, tumours completely
CC   regressed after 11 days.
SQ   Sequence   236 AA;


  Query Match             62.2%;  Score 790.5;  DB 1;  Length 236;
  Best Local Similarity   63.9%;  Pred. No. 7.5e-49;
  Matches  154;  Conservative   32;  Mismatches   46;  Indels    9;  Gaps    2;

Qy      1 MQLTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRF 60
          :||||||||:|||||||||||| |: ||| |||||||  :||| |:| || |||||| ||
Db      2 LQLTQSPAILSASPGEKVTMTCRATPSVSYMHWYQQKPGSSPKPWIYTTSNLASGVPARF 61

Qy     61 SGSGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIKGSTSGSGKSSEGKGV 120
          || ||| |||||:| :|||||||||| | |  | ||| |:||||||||||||||||||||
Db     62 SGGGSGTSYSLTVSRVEAEDAATYYCQQWSRSPPTFGGGSKLEIKGSTSGSGKSSEGKGV 121

Qy    121 QLQQSGVELVRGGALVKLSCKASDFNIKDYYIHWVKQRPEQGLEWIGWIHPENGNTVYDP 180
          ||| || |:|: |  :|:||| | ::   : ::||||   : ||||| |:| ||:| |:
Db    122 QLQDSGPEVVKPGGSMKISCKTSGYSFTGHTMNWVKQSHGKNLEWIGLINPYNGDTNYNQ 181

Qy    181 KFQGKASITADTSSNAAYLQLSSLTSEDTAVYYCA---SYYYYSAYYAMYYWGQGTSVTV 237
          ||:|||: | | ||: ||::| ||||||:|:||||   : :|:        || ||:|||
Db    182 KFKGKATFTVDKSSSTAYMELLSLTSEDSALYYCARRVTDWYFDV------WGAGTTVTV 235

Qy    238 S 238
          |
Db    236 S 236


RESULT    8
R15443
ID   R15443 standard; Protein; 240 AA.
AC   R15443;
DT   25-FEB-1992  (first entry)
DE   Single chain Fv from pSCV1 for blocking HRV binding to ICAM-1.
KW   antigen-binding fragment; inflammation; auto-immune disease.
OS   Homo sapiens.
PN   EP-459577-A.
PD   04-DEC-1991.
PF   25-MAY-1991; 201243.
PR   01-JUN-1990; US-532001.
PA   (MERI ) MERCK & CO INC.
PI   Colonno RJ, Condra JH, Tomassini JE, Sardana VV;
DR   WPI; 91-355850/49.
PT   Microbially expressed portions of monoclonal antibody - can block
PT   attachment of rhinovirus ligands to inter-cellular adhesion
PT   molecule (ICAM-1)
PS   Claim 5; Page 21; 28pp; English.
CC   The Fv fragment encoded by pSCV1 was able to protect cells against
CC   infection by HRV-14. It was less active than modified Fv fragments
CC   which included a Gly-Gly-Gly-Gly-Ser flexible linker between the L
CC   and H chains as either a monomer or a dimer.
CC   See also R15437-R15442.
SQ   Sequence   240 AA;


  Query Match             60.0%;  Score 763;  DB 1;  Length 240;
  Best Local Similarity   61.6%;  Pred. No. 6.5e-47;
  Matches  149;  Conservative   33;  Mismatches   50;  Indels   10;  Gaps    3;

Qy      3 LTQSPAIMSASPGEKVTMTCSASSSVS-NMHWYQQKSSTSPKLWVYDTSKLASGVPGRFS 61
          |||||| :| :||: |:::| || |:| |:|||||||  ||:| :   |:  ||:| |||
Db      4 LTQSPATLSVTPGDSVSLSCRASQSISNNLHWYQQKSHESPRLLIKHASQSISGIPSRFS 63

Qy     62 GSGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIKGSTSGSGKS----SEG 117
          |||||  ::|:|:|:| ||   ::| | : :|:||| ||||||| : :    |    |
Db     64 GSGSGTDFTLSINSVETEDFGMFFCQQSNSWPYTFGGGTKLEIKRADAAPTVSIFPPSSE 123

Qy    118 KGVQLQQSGVELVRGGALVKLSCKASDFNIKDYYIHWVKQRPEQGLEWIGWIHPENGNTV 177
          : ||||||| |||: || ||||| || ||||| ||||:|||||||||||| | | | ||:
Db    124 EEVQLQQSGAELVKPGASVKLSCTASGFNIKDTYIHWMKQRPEQGLEWIGRIDPANDNTI 183

Qy    178 YDPKFQGKASITADTSSNAAYLQLSSLTSEDTAVYYCASYYYYSAYYAMYYWGQGTSVTV 237
          |||| ||||::||||||| |||||||||||||||||| :  |: |     |||||| |||
Db    184 YDPKVQGKATMTADTSSNTAYLQLSSLTSEDTAVYYCTTSGYWFA-----YWGQGTLVTV 238

Qy    238 SS 239
          |:
Db    239 SA 240


RESULT    9
W25784
ID   W25784 standard; Protein; 240 AA.
AC   W25784;
DT   22-DEC-1997  (first entry)
DE   Anti-gp130 antibody derived scFv.
KW   Single chain Fv; scFv; antibody; E. coli; anti-T3 antibody;
KW   anti-gp130 antibody; inclusion body; chaperonin.
OS   Synthetic.
PN   J09220092-A.
PD   26-AUG-1997.
PF   15-FEB-1996; 027622.
PR   15-FEB-1996; JP-027622.
PA   (TOYJ ) TOSOH CORP.
DR   WPI; 97-474306/44.
DR   N-PSDB; T91615.
PT   Producing single chain Fv antibody in Escherichia coli - by
PT   expression in an inclusion body, followed by protein folding or by
PT   co-expression with a chaperonin as a soluble fraction
PS   Claim 4; Page 6-7; 9pp; Japanese.
CC   The sequences given in W25783-84 represent single chain Fv (scFv)
CC   antibodies which are produced in E.coli.  The scFV's are derived
CC   from an anti-T3 antibody or an anti-gp130 antibody and are produced
CC   either by: (1) expression as an inclusion body, followed by folding
CC   (i.e. by denaturation and solubilisation) or (2) expression as a
CC   soluble fraction by co-expression with chaperonin.  Using the methods,
CC   scFv can be produced in E. coli, either in a soluble fraction or in
CC   inclusion bodies.  Like the parent double chain antibodies, the scFv
CC   polypeptides are useful in immunological diagnosis or for separation
CC   and purification, but they lessen the side effects caused by antibody
CC   constant regions.  The availability of scFv polypeptides contributes
CC   to fundamental research and development of diagnostic and therapeutic
CC   drugs.
SQ   Sequence   240 AA;


  Query Match             59.7%;  Score 759;  DB 1;  Length 240;
  Best Local Similarity   63.2%;  Pred. No. 1.2e-46;
  Matches  151;  Conservative   31;  Mismatches   53;  Indels    4;  Gaps    3;

Qy        3 LTQSPAIMSASPGEKVTMTCSASSSVSNMHWYQQKSSTSPKLWVYDTSKLASGVPGRFSG 62
            ||||||::||||||||||||: ||||::|:|||||   :||| |:| || |||||| | ||
Db        4 LTQSPALISASPGEKVTMTCNVSSSVTSMYWYQQKPGSSPKPWIYLTSNLASGVPPRSSG 63

Qy       63 SGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIKGSTSGSGKS-SEGKGVQ 121
            |||| :|||||||:|||||||||| | |  | |||:|||||: |  || | |   |  ||
Db       64 SGSGTTYSLTISSLEAEDAATYYCQQWSTNPLTFGAGTKLELGGGGSGGGGSGGGGSDVQ 123

Qy      122 LQQSGVELVRGGALVKLSCKASDFNIKDYYIHWVKQRPEQGLEWIGWIHPENGNTVYDPK 181
            ||||| |||: || ||: |||| :   || : ||||    : ||||| |:  :|  :|: |
Db      124 LQQSGPELVKPGASVKIPCKASGYTFTDYNMDWVKQSHGKSLEWIGDINSHSGGIIYNQK 183

Qy      182 FQGKASITADTSSNAAYLQLSSLTSEDTAVYYCA-SYYYYSAYYAMYYWGQGTSVTVSS 239
            |: || :| | ||: ||::| ||||||||||||| :|| |  ::   ||||||::||||
Db      184 FKDKAILTVDKSSSTAYMELRSLTSEDTAVYYCARTYYVYGHFFD--YWGQGTTLTVSS 240


RESULT  10
W11508
ID   W11508 standard; Protein; 553 AA.
AC   W11508;
DT   24-SEP-1997  (first entry)
DE   Single chain anti-Fc gamma RI antibody fused to anti-CEA antibody
KW   Humanised antibody; anti-Fc receptor; H22; bifunctional; bispecific;
KW   fusion protein; chimera; carcinoembryonic antigen; CEA.
OS   Chimeric - Mus musculus.
OS   Chimeric - Homo sapiens.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   peptide         1. .19
FT                   /label= signal_peptide
FT   domain          20. .139
FT                   /label= H22_VH
FT                   /note= "Anti-Fc gamma RI antibody H22 heavy chain
FT                   variable domain"
FT   region          140. .154
FT                   /label= linker
FT   domain          155. .266
FT                   /label= H22_VL
FT                   /note= "Anti-Fc gamma RI antibody H22 light chain
FT                   variable domain"
FT   region          267. .276
FT                   /label= linker
FT   domain          277. .396
FT                   /label= MFE-23_VH
FT                   /note= "Anti-CEA antibody MFE-23 heavy chain variable
FT                   domain"
FT   region          397. .411
FT                   /label= linker
FT   domain          412. .520
FT                   /label= MFE-23_VL
FT                   /note= "Anti-CEA antibody MFE-23 light chain variable
FT                   domain"
FT   region          521. .533
FT                   /label= linker
```

```
FT   peptide         534. .544
FT                   /label= c-myc_tag
FT   region          548. .553
FT                   /label= His-6_tail
FT                   /note= "6 histidine residues"
PN   WO9640789-A1.
PD   19-DEC-1996.
PF   07-JUN-1996; U09988.
PR   07-JUN-1995; US-484172.
PA   (MEDA-) MEDAREX INC.
PI   Deo YM,  Goldstein J,  Graziano R,  Somasundaram C;
DR   WPI; 97-052242/05.
DR   N-PSDB; T58130.
PT   Recombinant, multi-specific anti-Fc receptor antibody molecules  -
PT   also comprise an anti-target portion, used for the treatment of
PT   cancer, autoimmune disease and pathogenic infection
PS   Example 8; Fig 40; 115pp; English.
CC   A mammalian expression construct encoding a single chain antibody
CC   having binding specificity for Fc gamma RI, derived from the
CC   humanised anti-Fc gamma RI monoclonal antibody H22 was prepared
CC   (see T58129). A bispecific single chain polypeptide was produced
CC   by fusing the H22 ScFv to an anti-carcinoembryonic antigen (CEA)
CC   antibody. The H22-anti-CEA fusion protein was shown to bind both
CC   Fc gamma RI and CEA.
SQ   Sequence   553 AA;


  Query Match             59.2%;  Score 752;  DB 1;  Length 553;
  Best Local Similarity   60.9%;  Pred. No. 9.4e-46;
  Matches  154;  Conservative   27;  Mismatches   46;  Indels   26;  Gaps    7;

Qy        1 MQLTQSPAIMSASPGEKVTMTCSASSSV---SN----MHWYQQKSSTSPKLWVYDTSKLA 53
            :||||||: :||| |::||:|| :| ||   ||    : |||||    :||| :|  |
Db      156 IQLTQSPSSLSASVGDRVTITCKSSQSVLYSSNQKNYLAWYQQKPGKAPKLLIYWASTRE 215

Qy       54 SGVPGRFSGSGSGNSYSLTISSMEAEDAATYYCYQGSGY--PFTFGSGTKLEIKGSTSGS 111
            |||| ||||||||  :: ||||:: || |||||:|   |   :||| |||:||| | |
Db      216 SGVPSRFSGSGSGTDFTFTISSLQPEDIATYYCHQ---YLSSWTFGQGTKVEIKSSCSSG 272

Qy      112 GKSSEGKGVQLQQSGVELVRGGALVKLSCKASDFNIKDYYIHWVKQRPEQGLEWIGWIHP 171
            |  |:   ::||||| |||| |  ||||| || ||||| |:||::| ||||||||||| |
Db      273 GGGSD---IKLQQSGAELVRSGTSVKLSCTASGFNIKDSYMHWLRQGPEQGLEWIGWIDP 329

Qy      172 ENGNTVYDPKFQGKASITADTSSNAAYLQLSSLTSEDTAVYYC-----ASYYYYSAYYAM 226
            |||:| | |||||||: | ||||| ||||||||||||||||||        ||:
Db      330 ENGDTEYAPKFQGKATFTTDTSSNTAYLQLSSLTSEDTAVYYCNEGTPTGPYYFD----- 385

Qy      227 YYWGQGTSVTVSS 239
             ||||||:|||||
Db      385 -YWGQGTTVTVSS 396


RESULT  11
W73223
ID   W73223 standard; Protein; 553 AA.
AC   W73223;
DT   25-JAN-1999  (first entry)
DE   H22-anti-CEA antibody construct 321.
KW   Multispecific single chain antibody; antibody H22; tumour cell; therapy;
KW   antibody-dependent cellular cytotoxicity; ADCC; HER 2/neu; infection;
KW   epidermal growth factor receptor; breast cancer; ovarian cancer.
OS   Synthetic.
PN   US5837243-A.
PD   17-NOV-1998.
PF   07-JUN-1996; 661052.
PR   07-JUN-1996; US-661052.
PR   07-JUN-1995; US-484172.
PA   (MEDA-) MEDAREX INC.
PI   Deo YM, Goldstein J, Graziano R, Somasundaram C;
DR   WPI; 99-023374/02.
DR   N-PSDB; V08176.
PT   Specific killing of tumour cells - using a multi-specific molecule
PT   comprising an anti-Fc receptor antibody and a portion which binds to
PT   a target cell
PS   Example 8; Fig 40; 57pp; English.
CC   This sequence represents the construct 321, which is a fusion between an
CC   anti-CEA antibody and a multispecific single chain antibody designated
CC   H22. The antibody can be used in the method of the invention for inducing
CC   antibody-dependent cellular cytotoxicity (ADCC) against a tumour cell
CC   which is characterised by overexpression of HER 2/neu or epidermal growth
CC   factor receptor (EGFR), comprises contacting the tumour cell with a
CC   multispecific protein molecule (preferably a single chain antibody)
CC   comprising: (a) an anti-Fc receptor antibody or an antigen binding
CC   fragment; (b) a portion which binds to HER 2/neu; and (c) a portion which
CC   binds to EGFR. The method can be used for treating cancers especially
CC   breast cancer or ovarian cancer. The multispecific antibody can also
CC   be administered prophylactically to vaccinate a subject against infection
CC   by a target cell.
SQ   Sequence   553 AA;

  Query Match             59.2%;  Score 752;  DB 1;  Length 553;
  Best Local Similarity   60.9%;  Pred. No. 9.4e-46;
  Matches  154;  Conservative   27;  Mismatches   46;  Indels   26;  Gaps    7;

Qy        1 MQLTQSPAIMSASPGEKVTMTCSASSSV---SN----MHWYQQKSSTSPKLWVYDTSKLA 53
            :||||||: :||| |::||:|| :| ||   ||    : |||||   :||| :|  |
Db      156 IQLTQSPSSLSASVGDRVTITCKSSQSVLYSSNQKNYLAWYQQKPGKAPKLLIYWASTRE 215

Qy       54 SGVPGRFSGSGSGNSYSLTISSMEAEDAATYYCYQGSGY--PFTFGSGTKLEIKGSTSGS 111
            |||| ||||||||  :: ||||:: || |||||:|   |   :||| |||:||| | |
Db      216 SGVPSRFSGSGSGTDFTFTISSLQPEDIATYYCHQ---YLSSWTFGQGTKVEIKSSCSSG 272

Qy      112 GKSSEGKGVQLQQSGVELVRGGALVKLSCKASDFNIKDYYIHWVKQRPEQGLEWIGWIHP 171
            |  |:   ::||||| |||| |  ||||| || ||||| |:||::| ||||||||||| |
Db      273 GGGSD---IKLQQSGAELVRSGTSVKLSCTASGFNIKDSYMHWLRQGPEQGLEWIGWIDP 329

Qy      172 ENGNTVYDPKFQGKASITADTSSNAAYLQLSSLTSEDTAVYYC-----ASYYYYSAYYAM 226
            |||:| | |||||||: | ||||| ||||||||||||||||||       ||:
Db      330 ENGDTEYAPKFQGKATFTTDTSSNTAYLQLSSLTSEDTAVYYCNEGTPTGPYYFD----- 385

Qy      227 YYWGQGTSVTVSS 239
             ||||||:|||||
Db      385 -YWGQGTTVTVSS 396

RESULT  12
W02191
ID   W02191 standard; Protein; 239 AA.
AC   W02191;
DT   13-NOV-1996  (first entry)
DE   18-2-3/TRY202' single chain binding protein.
KW   Antibody engineering; monoclonal antibody; MAb;
KW   single chain antibody; immunoassay; fluorescein; 18-2-3/TRY202';
KW   single chain binding protein.
OS   Chimeric Mus musculus;
OS   Chimeric synthetic.
PN   US5518889-A.
PD   21-MAY-1996.
PF   02-SEP-1986; 902971.
PR   02-SEP-1986; US-902971.
PR   02-SEP-1987; US-092110.
PR   19-JAN-1989; US-299617.
PR   25-APR-1990; US-512910.
PR   01-APR-1993; US-040440.
PR   06-JUN-1995; US-468988.
PA   (ENZO-) ENZON LABS INC.
PI   Bird RE,  Hardman K,  Ladner RC;
DR   WPI; 96-259060/26.
DR   N-PSDB; T36463.
PT   Immunoassay using single chain antigen binding mol. - as replacement
PT   for labelled or immobilised antibody, are less immunogenic, easier
PT   to engineer, more stable and less expensive
PS   Example 9; Fig 39A-B; 78pp; English.
CC   Computer-designed construct 18-2-3/TRY202' (W02191), encoded by
CC   the DNA sequence given in T36463, comprises the variable regions
```

```
CC   of the light and heavy chains of anti-fluorescein monoclonal
CC   antibody (Mab) 18-2-3 linked by a peptide designed to fit into a
CC   groove on the backside of the variable domain structure.  The DNA
CC   construct was inserted into vector pGX3703 and introduced into E.
CC   coli.  18-2-3/TRY202' was produced as a single chain molecule.  It
CC   exhibited biological binding activity equivalent in specifically
CC   and affinity to that of the original MAb.
SQ   Sequence   239 AA;

  Query Match             59.1%;  Score 751.5;  DB 1;  Length 239;
  Best Local Similarity   62.2%;  Pred. No. 4.1e-46;
  Matches  148;  Conservative   30;  Mismatches   55;  Indels    5;  Gaps    3;

Qy        3 LTQSPAIMSASPGEKVTMTCSASSSVSN--MHWYQQKSSTSPKLWVYDTSKLASGVPGRF 60
            |||||||||||||||||||| ||||||:  :|||||||  ||||||| || |||||| ||
Db        5 LTQSPAIMSASPGEKVTMTCRASSSVSSSYLHWYQQKSGASPKLWVYGTSNLASGVPARF 64

Qy       61 SGSGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIKGSTSGSGKSSEGKGV 120
            || ||| ||||||||:|||||||||| | |||| |||:|||||::| :||||  || |
Db       65 SGGGSGTSYSLTISSVEAEDAATYYCQQYSGYPLTFGAGTKLELEGKSSGSG--SESKST 122

Qy      121 QLQQSGVELVRGGALVKLSCKASDFNIKDYYIHWVKQRPEQGLEWIGWIHPENGNTVYDP 180
            ||::||  ||     : ::|  | |:: :| :|||:| | :||||:| |    ||| |:
Db      123 QLKESGPVLVAPSQSLSITCTVSGFSLTNYGVHWVRQPPGKGLEWLGVIW-AGGNTNYNS 181

Qy      181 KFQGKASITADTSSNAAYLQLSSLTSEDTAVYYCASYYYYSAYYAMYYWGQGTSVTVS 238
                : ||: | | :  :|:::|| ::|||:||||       |||| |||||||||||
Db      182 ALMSRLSISKDNSKSQVFLKMNSLQNDDTAIYYCAKRLGRIFYYAMDYWGQGTSVTVS 239

RESULT  13
R43679
ID   R43679 standard; Protein; 239 AA.
AC   R43679;
DT   23-MAY-1994  (first entry)
DE   Single chain polypeptide with affinity for fluorescein.
KW   Monoclonal antibody; MAb; affinity; binding; antigen; diagnostics;
KW   therapy; imaging; purification; biosensors.
OS   Synthetic.
PN   US5260203-A.
PD   09-NOV-1993.
PF   02-SEP-1986; 902971.
PR   02-SEP-1986; US-902971.
PR   02-SEP-1987; US-092110.
PR   19-JAN-1989; US-299617.
PR   25-APR-1990; US-512910.
PA   (ENZO-) ENZON INC.
PI   Bird RE,  Hardman K;  Ladner RC;
DR   WPI; 93-367875/46.
DR   N-PSDB; Q51540.
PT   Single chain poly:peptide for binding antigen - comprising light
PT   and heavy chain antigen binding portions linked by peptide linker
PS   Example 9; Figure 39; 78pp; English.
CC   The single chain polypeptide is derived from the mature light and
CC   heavy chains of a monoclonal antibody (MAb) and has affinity
CC   for a given antigen (Fluorescein).  It comprises a first
CC   polypeptide comprising the antigen binding portion of of the light
CC   chain variable region of an antibody and a second polypeptide
CC   comprising the antigen binding portion of the heavy chain variable
CC   region of an antibody and at least one peptide linker linking the
CC   first and second polypeptide chains.  The resulting single chain
CC   polypeptide can be used in diagnostics, therapy
CC   (in vivo and in vitro), imaging, purifications and biosensors.
CC   This particular single chain binding molecule was designated
CC   18-2-3/TRY202' and contains one linker peptide.
SQ   Sequence   239 AA;

  Query Match             58.9%;  Score 748.5;  DB 1;  Length 239;
  Best Local Similarity   62.2%;  Pred. No. 6.7e-46;
  Matches  148;  Conservative   29;  Mismatches   56;  Indels    5;  Gaps    3;
```

```
Qy      3 LTQSPAIMSASPGEKVTMTCSASSSVSN--MHWYQQKSSTSPKLWVYDTSKLASGVPGRF 60
          |||||||||||||||||||| ||||||:  :|||||||  ||||||| || |||||| ||
Db      5 LTQSPAIMSASPGEKVTMTCRASSSVSSSYLHWYQQKSGASPKLWVYGTSNLASGVPARF 64

Qy     61 SGSGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIKGSTSGSGKSSEGKGV 120
          || ||| ||||||||:|||||||||| | |||| |||:|||||::| :||||  || |
Db     65 SGGGSGTSYSLTISSVEAEDAATYYCQQYSGYPLTFGAGTKLELEGKSSGSG--SESKST 122

Qy    121 QLQQSGVELVRGGALVKLSCKASDFNIKDYYIHWVKQRPEQGLEWIGWIHPENGNTVYDP 180
          ||::||  ||     : ::|  | |:: :| :|||:| | :||||:| |    ||| |:
Db    123 QLKESGPVLVAPSQSLSITCTVSGFSLTNYGVHWVRQPPGKGLEWLGVIW-AGGNTNYNS 181

Qy    181 KFQGKASITADTSSNAAYLQLSSLTSEDTAVYYCASYYYYSAYYAMYYWGQGTSVTVS 238
              : ||: | | :  :|:::||  :|||:||||       |||| |||||||||||
Db    182 ALMSRLSISKDNSKSQVFLKMNSLQIDDTAIYYCAKRLGRIFYYAMDYWGQGTSVTVS 239


RESULT  14
R99649
ID   R99649 standard; Protein; 239 AA.
AC   R99649;
DT   11-OCT-1996  (first entry)
DE   Single chain binding molecule 18-2-3/TRY202'.
KW   Antibody engineering; single polypeptide chain binding molecule;
KW   single chain antibody; SCA; heavy chain; light chain;
KW   monoclonal antibody; MAb; immunoaffinity purification;
KW   18-2-3/TRY202'.
OS   Chimeric Mus sp;
OS   Chimeric synthetic.
PN   US5534621-A.
PD   09-JUL-1996.
PF   02-SEP-1986; 902971.
PR   02-SEP-1986; US-902971.
PR   02-SEP-1987; US-092110.
PR   19-JAN-1989; US-299617.
PR   25-APR-1990; US-512910.
PR   01-APR-1993; US-040440.
PR   06-JUN-1995; US-468992.
PA   (ENZO-) ENZON LABS INC.
PI   Bird RE,  Hardman K,  Ladner RC;
DR   WPI; 96-333309/33.
DR   N-PSDB; T13739.
PT   Immuno:purificn. using single binding chain molecule including
PT   antigen-binding parts of antibody light and heavy chain variable
PT   regions connected by a linker - is smaller, stabler and less
PT   expensive than complete antibodies
PS   Example 9; Fig 39; 78pp; English.
CC   A DNA construct (T13739) codes for single chain binding molecule
CC   18-2-3/TRY202' (R99649), in which VL and VH regions of anti-
CC   fluorescein monoclonal antibody 18-2-3 are joined by a peptide
CC   linker composed primarily of alternating Gly and Ser residues, with
CC   Glu and Lys residues to enhance solubility.  The construct was
CC   expressed in E.coli.  Soluble, folded chain antibody (SCA) was
CC   obtd. which was capable of exhibiting a biological binding activity
CC   equivalent in specificity and affinity to that of a monoclonal
CC   antibody.  SCAs (see also R99645-48 and T99650-51) have the
CC   advantages of smaller size, greater stability and reduced cost.
SQ   Sequence   239 AA;


  Query Match             58.9%;  Score 748.5;  DB 1;  Length 239;
  Best Local Similarity   62.2%;  Pred. No. 6.7e-46;
  Matches  148;  Conservative   29;  Mismatches   56;  Indels    5;  Gaps    3;

Qy      3 LTQSPAIMSASPGEKVTMTCSASSSVSN--MHWYQQKSSTSPKLWVYDTSKLASGVPGRF 60
          |||||||||||||||||||| ||||||:  :|||||||  ||||||| || |||||| ||
Db      5 LTQSPAIMSASPGEKVTMTCRASSSVSSSYLHWYQQKSGASPKLWVYGTSNLASGVPARF 64

Qy     61 SGSGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIKGSTSGSGKSSEGKGV 120
          || ||| ||||||||:|||||||||| | |||| |||:|||||::| :||||  || |
Db     65 SGGGSGTSYSLTISSVEAEDAATYYCQQYSGYPLTFGAGTKLELEGKSSGSG--SESKST 122

Qy    121 QLQQSGVELVRGGALVKLSCKASDFNIKDYYIHWVKQRPEQGLEWIGWIHPENGNTVYDP 180
          ||::||  ||     : ::|  | |:: :| :|||:| | :||||:| |    ||| |:
Db    123 QLKESGPVLVAPSQSLSITCTVSGFSLTNYGVHWVRQPPGKGLEWLGVIW-AGGNTNYNS 181

Qy    181 KFQGKASITADTSSNAAYLQLSSLTSEDTAVYYCASYYYYSAYYAMYYWGQGTSVTVS 238
              : ||: | | :  :|:::||  :|||:||||       |||| |||||||||||
Db    182 ALMSRLSISKDNSKSQVFLKMNSLQIDDTAIYYCAKRLGRIFYYAMDYWGQGTSVTVS 239


RESULT  15
R06482
ID   R06482 standard; protein; 241 AA.
AC   R06482;
DT   07-JAN-1991  (first entry)
DE   18-2-3-/TRY202'.
KW   Monoclonal antibody; variable chain; anti-fluorescein.
PN   US4946778-A.
PD   07-AUG-1990.
PF   19-JAN-1989; 299617.
PR   02-SEP-1986; US-902971.
PR   02-SEP-1987; US-092110.
PR   19-JAN-1989; US-299617.
PA   (GENE-) GENEX CORP.
PI   Ladner RC, Bird RE, Hardman K;
DR   WPI; 90-260350/34.
DR   P-PSDB; R06482.
PT   Single polypeptide chain binding molecules - having light chain
PT   variable region of antibody linked by peptide to heavy chain
PT   variable region.
PS   Disclosure; Fig 39; 68pp; English.
CC   The sequence was constructed from the variable regions of an anti-
CC   fluorescein  MAb, 18-3-2, an IgM.   The VL and VH cDNA sequences
CC   were synthesised by priming on RNA isolated from hybridoma cells.
CC   The sequence encodes a single chain binding molecule comprising
CC   the variable regions of heavy and light chains linked by peptides.
CC   The peptide linker was computer-designed to fit inyo a groove on
CC   the backside of the variable domain structure and is composed
CC   primarly of alternating Gly and Ser residues. Glu and Lys residues
CC   are included  to enhance solubility.
CC   See also R06476-R06484.
SQ   Sequence   241 AA;


  Query Match             58.6%;  Score 744.5;  DB 1;  Length 241;
  Best Local Similarity   61.8%;  Pred. No. 1.3e-45;
  Matches  147;  Conservative   29;  Mismatches   57;  Indels    5;  Gaps    3;

Qy      3 LTQSPAIMSASPGEKVTMTCSASSSVSN--MHWYQQKSSTSPKLWVYDTSKLASGVPGRF 60
          |||||||||||||||| ||| ||||||:  :|||||||  ||||||| || |||||| ||
Db      5 LTQSPAIMSASPGEKVNMTCRASSSVSSSYLHWYQQKSGASPKLWVYGTSNLASGVPARF 64

Qy     61 SGSGSGNSYSLTISSMEAEDAATYYCYQGSGYPFTFGSGTKLEIKGSTSGSGKSSEGKGV 120
          || ||| ||||||||:|||||||||| | |||| |||:|||||::| :||||  || |
Db     65 SGGGSGTSYSLTISSVEAEDAATYYCQQYSGYPLTFGAGTKLELEGKSSGSG--SESKST 122

Qy    121 QLQQSGVELVRGGALVKLSCKASDFNIKDYYIHWVKQRPEQGLEWIGWIHPENGNTVYDP 180
          ||::||  ||     : ::|  | |:: :| :|||:| | :||||:| |    ||| |:
Db    123 QLKESGPVLVAPSQSLSITCTVSGFSLTNYGVHWVRQPPGKGLEWLGVIW-AGGNTNYNS 181

Qy    181 KFQGKASITADTSSNAAYLQLSSLTSEDTAVYYCASYYYYSAYYAMYYWGQGTSVTVS 238
              : ||: | | :  :|:::||  :|||:||||       |||| |||||||||||
Db    182 ALMSRLSISKDNSKSQVFLKMNSLQIDDTAIYYCAKRLERIFYYAMDYWGQGTSVTVS 239


Search completed: May 13, 1999, 10:02:42
Job time: 1500 sec
```